Next-Generation Railway Inspection Approaches Based on a Combination of Multiple Sensing Technologies

by

JIAQI YE

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Department of Electronic, Electrical
and Systems Engineering
School of Engineering
University of Birmingham
January 2022

UNIVERSITY OF BIRMINGHAM

Abstract

The railway has played a vital role in public transportation and has also been identified as the safest mode of land transport for decades. Key to this is the efficient and constantly improving safety management of some of its major subsystems. Among these, the inspection of railway tracks is an essential part. It ensures that the quality of the rails in operation meets relevant standards and supports appropriate remedial actions to prevent further deterioration or accidents. Rail inspection covers both internal and external defects. External defects, also known as surface defects, are normally derived from long-term rolling contact. They are more common and normally are the point of origin of more severe internal defects. Therefore, the inspection of surface defects is always a significant concern to the industry which should aim to make such inspections more accurate and efficient. In recent years, the development of advanced sensory technologies has automated a lot of rail inspection tasks, which has considerably improved the performance of rail inspection.

This research firstly introduces and categorises common rail surface defects under related standards and reviews the basic principles and performance of existing inspection technologies. The initiation and visible nature of surface defects determine that their inspection mainly involves the visual identification of rail faults and rail profile measurements. Therefore, laser- and vision-based inspection approaches are among the most popular solutions for the inspection of surface defects. However, as two optical sensing technologies, some of their inherent shortcomings can affect their effectiveness and limit their applications in specific scenarios. One of the potential solutions is to integrate multiple sensors, through which the strengths of different sensors can be combined to eliminate their shortcomings.

Accordingly, the second part of this research further analyses the strengths and shortcomings of laser and vision sensors for rail inspection through two practical case studies. This drives the design of a new prototype for a laser-based multi-sensing system. The new prototype uses a laser triangulation sensor as the profiling unit and integrates a camera and an inertial measurement unit (IMU) to assist the measurement, providing a more autonomous and portable device. Such a design not only allows for greater integrity of 2D rail profile measurement through an optimised iterative closest point (ICP) algorithm but also allows for 3D rail surface measurement and

reconstruction through a dedicated 2D profile alignment method. These further allow for rail inspection in multiple dimensions. The performance of the multi-sensing system in 2D rail profile measurement was tested under the scope of the S-CODE project. The feasibility of 3D measurement was also demonstrated through extensive lab and field trials.

The integration of multiple sensors not only offers more accurate and higher-dimensional measurement capabilities. The cross-over of inspection strategies contains more opportunities. The third part of this research analyses the transferability of processing approaches between 3D laser sensor data and 2D images. The state-of-the-art concepts of machine learning and deep learning in computer vision were then applied to processing high-precision 3D measurement data from the proposed system. These delivered two optimised solutions for inspecting surface defects: 1) end-to-end 3D pixel-level defect detection and 2) two-steps 3D model-based defect classification and evaluation. The testing results demonstrated that the first solution contributes to more precise defect detection and characterisation, and the second solution contributes to more comprehensive defect evaluation and classification. Both of them outperform most existing inspection strategies that rely on a single data form and show the potential to replace human inspectors to carry out rail surface defect inspection more accurately and efficiently.

By considering the performance of the individual inspection technologies alongside the increased capability of the integrated muti-sensing solution, and through the incorporation of intelligent algorithms that apply complex techniques to the newly available integrated data stream, this thesis demonstrates that an integrated approach provides improved inspection performance.

Acknowledgements

The four years' PhD research is a long and enjoyable journey that needs passion, insistence, and patience, and also cannot be done without support from others. I would like to thank the Birmingham Centre for Railway Research and Education (BCRRE) and China Scholarship Council for providing me with sponsorship. I would also like to take this chance to express my sincere thanks to those who have helped me during this journey.

Firstly, I would like to thank my supervisor, Dr Edward Stewart, for his excellent support and supervision throughout my PhD. We have known each other since the very beginning of my study at the University of Birmingham. It is his experienced guidance, friendship, and vision that have directed me to go through all challenges and become a qualified researcher.

I would like to thank Prof. Clive Roberts, for providing me with the opportunity to be a part of the BCRRE. Also, as my supervisor, he helped me choose my research topic and has provided me with professional and prospective guidance on my research.

I would like to thank my academic advisor Dr Stuart Hillmansen, for his constructive suggestions in the key stages of my PhD research.

I would like to thank Dr Lei Chen, for his practical advice on my research and for helping me to link part of my research with industrial projects.

I would like to thank all my colleagues at BCRRE. It is them that form a warm and supportive working environment. Particular thanks go to Mr Adnan Zentani and Mrs Sarah Yeo for their assistance with all the lab work; and to Dr Karthik Thangaraj, Dr Krishnan Venkateswaran, and Dr Graeme Yeo, for their kind help and support on the S-CODE project, in which I participated during my PhD.

Finally, I would like to thank my parents and my partner, Dr Qianyu Chen, for their love, support, and encouragement throughout my research course. In particular, I would like to share this moment with my grandma and express how much I love and miss her.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Term	Explanation/Meaning/Definition
1D/2D/3D	One-dimensional/two-dimensional/three-dimensional
AC	Alternating Current
ACFM	Alternating Current Field Measurement
BR	Boundary Refinement
CCD	Charge-Coupled Device
CHT	Circular Hough Transform
CMOS	Complementary Metal-Oxide Semiconductor
CNN	Convolutional Neural Network
DCS	Double Circle Segment
DoF	Degrees of Freedom
EMAT	Electromagnetic Acoustic Transducer
FN	False Negative
FCN	Fully Convolutional Network
FGR	Fast Global Registration
FNN	Feed Forward Neural Network
FP	False Positive
FRA	Federal Railroad Administration
GCA	Global Context Attention
GCC	Gauge Corner Cracking
GMM	Gaussian Mixture Model
GPS	Global Position System
GUI	Graphical User Interface
HD	High-Speed Digital
HOG	Histogram of Oriented Gradient
HSV	Hue, Saturation, Value
ICP	Iterative Closest Point
IMU	Inertial Measurement Unit
IoU	Intersection over Union
IR	Infrared
K-d tree	K-dimensional tree
KNN	K-nearest neighbour
LCD	Liquid Crystal Display
MIoU	Mean Intersection over Union

Term	Explanation/Meaning/Definition
ML/DL	Machine Learning/Deep Learning
MRF	Markov Random Field
MSER	Maximally Stable Extremal Regions
MVM	Majority Voting Mechanism
NDT	Non-Destructive Techniques
NMT	Network Rail's New Measurement Train
NR	Network Rail
PLPR	Plain Line Pattern Recognition
PMS	Profile Measurement System
RCF	Rolling Contact Fatigue
RGB	Red, Green Blue Colour System
RLU	Rectified Linear Unit
RMSE	Root-Mean-Square-Error
ROI	Region of Interest
RPN	Region Proposal Network
RSICP	Reweight-Scaling Iterative Closest Point
RVDS	Railway Visual Detection System
S&C	Switches & Crossings
S-CODE	Switch and Crossing Optimal Design and Evaluation
SGD	Stochastic Gradient Descent
SICP	Scaling Iterative Closest Point
SIFT	Scale-Invariant Feature Transform
SLAM	Simultaneous Localisation and Mapping
SMS	Surface Measurement System
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
UGM	Unattended (track), Geometry Measurement System
UIC	International Union of Railways
TP	Ture Positive
YOLO	You Only Look Once

1 INTRODUCTION

1.1 Background

Rail, originated in the UK, has played a vital role in public transportation for hundreds of years. After entering the 21st century, the rapid development of electrified and high-speed rail around the world has significantly increased the volume of railway transportation and boosted the world economy, which has again enhanced the significance of the railway as a transport backbone. According to statistics from the Department for Transport (DfT), rail usage has almost doubled in the last two decades in the UK; the number of rail passengers has increased by 97%, reaching a record of 1.8 billion in 2018/2019 [1]. In China, the development of railways following the country's 'reform and opening up' has increased to over 120,000 km including 35,000 km of high-speed rail in the last two decades, which has become a crucial means of transportation. Currently, rail accounts for 8% of the world's passenger transport and 7% of freight transport. These proportions are still rising steadily all over the world [2].

Against this background of rapid development, rail is still the safest mode of land transport. This cannot be separated away from efficient and constantly improving safety management of the railway system, which covers various subsystems including rolling stock, infrastructure, control, communication and signalling, information technology, and its human resources, etc. Among these, the inspection of railway tracks takes a large proportion of manpower and resources. As railway tracks are a major infrastructure in railway systems, their normal operation directly affects the safety and performance of the railway network. Normally, the inspection of railway tracks refers to the process of ensuring that the quality of the rails in operation meets relevant standards and that rail faults can be identified in their early stages, based on which appropriate maintenance works can be applied to prevent further deterioration and accidents. For a long time, human inspectors were the main actors for rail inspection; however, they can be inefficient and subjective with the involvement of human factors, and thus incompatible with the rapid expansion of the railway network and the ever-shortening maintenance windows available in modern railway systems. Therefore, new rail inspection concepts and approaches have always attracted the attention of both researchers and railway operators aiming to improve the performance of rail inspection. In recent years, a lot of new technologies have been developed and applied, for example

the use of ultrasonic, magnetic, eddy current, visual, and laser technologies, etc. These methods are mainly based on advanced sensory and measuring technologies, which measure rail conditions digitally or generate digital signals that can represent rail conditions. By exploiting the data through specific fault diagnosis algorithms, it is possible to identify different categories of rail faults such as surface defects, sub-surface defects, internal defects, and deformation more efficiently.

Overall, the application of these new technologies has considerably improved the performance of rail inspection. Along with the development of related sensory technologies, it can be expected that rail inspection is developing towards digitisation, automation, and intelligence, which ensures the optimal health status of rail in a more accurate, efficient, and reliable way with reduced human involvement.

1.2 Problem descriptions and trends

The tendency towards digitisation, automation, and intelligence makes most of the inspection strategies data-driven, which means that the quality of data generated from sensors is the prerequisite for further inspection works and determines the capability and effect of inspection directly. This brings two main challenges: 1) every single sensory technology inevitably has its weaknesses or inherent shortcomings in some specific application scenarios, which can influence its performance. For example, the eddy current technique utilises electromagnetic effects between the alternating current and the magnetic field to detect surface and near-surface defects. However, electromagnetic effects can be easily affected by interference signals derived from undesired factors such as the conductivity of the rail and rail welds; 2) the information required to represent a rail fault is normally multidimensional, including the location, size, and shape of defects, etc. However, the information that a single unit can provide is limited. For example, it is possible to detect the appearance of surface defects by using two-dimensional (2D) rail profiles from laser sensors, but further evaluation works such as categorisation and analysis of the severity, which requires a more comprehensive review of defects, still rely on human experts.

The capacity and information provided by an individual sensory unit normally cannot satisfy the requirements of a practical inspection task, even for faults of the same category. Therefore, most state-of-the-art rail inspection systems integrate different sensors, through which the strengths of different sensors can be combined to eliminate

their shortcomings. For example, the robotic trolley developed by Hamed Rowshandel uses a combination of alternating current field measurement, a laser triangulation sensor, and a robotic arm to allow fully automated detection and characterisation of rail surface defects [3]. In industry, a representative system is Network Rail's New Measurement Train, which is equipped with 14 sensors including high-precision laser scanners, high-resolution cameras, and accelerometers, etc. The multi-sensing data can be fed into onboard analysis systems to comprehensively reflect rail conditions and help identify faults [4].

In the context of digitisation, the integration of multiple sensors, also known as sensor fusion, is another important trend for future rail inspection approaches, which enhances the capability of inspection systems and allows comprehensive fault diagnosis with more informatic and reliable multi-sensing data available.

1.3 Scope and terminology

The term 'rail' in this research typically refers to the rail track in the railway system. Normally, the inspection of rails covers both internal and external defects. This research has been focused on the inspection of external defects since they are more common and normally the origin of more severe internal defects. External defects can be further divided into three groups based on their origin: (i) manufacturing defects, (ii) defects caused by improper handling, use, and installation, and (iii) defects as a result of rolling contact fatigue. According to the investigation by Cannon et al. [5], defects belonging to the first two groups have been dramatically reduced with the use of advanced alloys and the improvement of rail manufacturing industries, while the defects of the third group are difficult to control as they can originate from the long-term rolling contact between the rail and train wheels. The cause of these defects determines that most of them are visible and distributed on the railhead surface where the train wheels make contact, and thus conventional manual inspections mainly involve visual identification of rail faults and rail profile measurements. Therefore, laser- and vision-based inspection approaches are among the most popular solutions to replace manual operations to automate the process of railhead surface defect inspection. For rail inspection, laser-based technologies mainly refer to the use of laser triangulation sensors which measure the rail geometry (e.g., rail profile) and physical surface conditions based on optical reflections; vision technologies mainly rely on the use of

2D imaging systems, such as HD cameras, to represent rail conditions through 2D images or video frames.

From the image-forming perspective, both laser and vision sensors are based on optical reflections. The only difference is that laser sensors receive active optical reflections by projecting structured laser light onto the target surface, while vision sensors capture passive nature light reflections. The fundamental point of optical reflections determines that both laser and vision sensors are truly non-contact. Compared to conventional full-contact (e.g., mechanical gauges) or semi-contact instruments (e.g., ultrasonic systems where a medium is required between the transducers and the target surface), they are naturally more efficient. Therefore, laser- and vision-based inspection systems have increasingly been developed and applied for rail inspection and have also been the main research objects in this research. Currently, most inspection systems are operated with such optical sensors mounted on trains [4] [6], rail vehicles [7] [8], or trolleys [9] [10]. The application of these onboard systems has greatly improved the efficiency of rail inspection and reduced the dependency on manual inspections. However, as mentioned in section 1.2, the performance of an individual sensory unit can be affected by its inherent shortcomings and external factors in specific areas. This also commonly occurs in current laser- and vision-based rail inspection systems. Some distinct shortcomings of laser and visual sensors are summarised below; they will be expanded and introduced in detail in the discussion of the literature in Chapter 2 and the case studies in Chapters 3 and 4.

- The shadowing effects of laser sensors.

- Additional fixture or stability required for laser sensors.

- The low robustness of visual sensors to surface conditions and external lights.

For rail inspection, these inherent shortcomings of sensors can result in other problems, for example incomplete rail profile measurement due to the shadowing effects of laser sensors, the bulky size and limited degrees of freedom (DoF) of the system due to the dependence on additional mechanical frames of laser sensors, and the low accuracy of visual sensors in practice due to the difficulty of maintaining the cleanliness of the rail surface. Moreover, existing laser and vision systems for rail inspection are normally operated on a 2D data basis. As a result, a lot of inspection tasks that require higher-dimensional information still rely on human experts to carry out detailed investigations.

Therefore, whether from the perspective of the system, measurement data, or specific inspection algorithms, there are still some gaps to fill to improve the performance of laser- and vision-based rail inspection approaches.

1.4 Research aims and objectives

Given the existing gaps in rail inspection approaches based on optical sensors (laser and vision) and the trend of integrating multiple sensors, a hypothesis was proposed to identify whether it is possible to improve the overall performance of laser- and vision-based rail inspection through the combination of multiple sensors. The research works that come with the hypothesis can be categorised into three aspects: (1) identification of the characteristics and existing gaps of laser- and vision-based rail inspection, (2) design of a new prototype for a laser-and vision-based multi-sensing system, and (3) development of intelligent inspection algorithms based on multi-sensing data.

The first part of this research aims to clarify the requirements for rail inspection in accordance with related standards and to identify existing gaps. These mainly include the identification of inspection targets, the categorisation of defects, and a review of existing inspection methods. Among existing inspection methods, vision- and laser-based inspection methods are reviewed in detail to identify their existing gaps for rail inspection. Relevant multi-sensor technologies are reviewed to help explore potential ways to integrate different sensors as one system for rail inspection.

Specific works and objectives that come with the first part of this research can be summarised as follows:

- to identify specific categories of rail defects for this research and the inspection requirements in accordance with related inspection standards.

- to identify the fundamentals of existing inspection solutions and their structural characteristics.

- to investigate existing gaps of vision- and laser-based inspection systems when facing specific rail inspection tasks.

- to initially identify the complementarity between vision- and laser-based inspection technologies and the feasibility of combining multiple sensors.

The second part of this research aims to design and develop a new prototype for a laser- and vision-based multi-sensing system for rail inspection. The term 'multi-sensing' in

this research refers to the combination of multiple sensors, which involves both the integration of multiple sensors at the hardware level and the fusion of multi-sensing data at the data level. Associated works from this part were also in line with the S-CODE (Switch and Crossing Optimal Design and Evaluation) project under the European Union's Horizon 2020 research and innovation programme, which included a work package aimed to design a more autonomous, robust, and miniaturised multi-sensing prototype for rail measurement and inspection [11]. This ensures the practicability of the proposed multi-sensing system and ascertains that its design satisfies industrial requirements.

Specific works and objectives that come with the second part of this research can be summarised as follows:

- to further analyse the characteristics of vision- and laser-based rail inspection technologies through two case studies, to identify their complementary strengths for the potential rail inspection tasks within the framework of this research.

- to design a new prototype of a multi-sensing system that fully considers the characteristics of laser and vision sensors and combines their strengths for rail inspection, to mitigate existing gaps from the hardware point of view.

- to develop dedicated data fusion algorithms to fuse multi-sensing data to help adapt the system to specific rail inspection tasks and to measure and characterise rail conditions more accurately and comprehensively.

With more accurate and higher-dimensional measurement data from the multi-sensing system, more useful information can be extracted for rail inspection. This gives the chance to develop more intelligent railway-specific inspection algorithms. Conventionally, rail inspection algorithms are mainly based on human-dominated parameters or thresholds settings, and handcrafted features to distinguish and describe the patterns corresponding to specific rail faults. These methods are, therefore, subjective, heavily reliant on human experience, and difficult to generalise for diverse cases in the real world. The recent development of machine learning methods allows a machine to learn feature representations from raw data independently, which helps to map input data to the expected output targets directly. With sufficient data, these machine learning models can be trained on purpose to carry out a lot of inspection tasks

that were originally dominated by humans. Accordingly, the third part of this research aims to develop more intelligent rail inspection algorithms using multi-sensing data, to reduce the dependency on human experts.

Specific works and objectives that correspond to the third part of this research are summarised as follows:

- to identify the complementarity between vision- and laser-based inspection algorithms and their applicability for the multi-sensing data.

- to identify patterns that can be used to describe specific rail faults from the multi-sensing data.

- to replace conventional human-dominated inspection methods with automatic inspection models to carry out rail inspection more accurately and efficiently.

1.5 List of publications

Publications from this research are listed below for reference.

1. **J. Ye**, E. Stewart, C. Roberts, et al., "Laser-based end-to-end 3D pixel-level rail surface defect detection and characterization method," in *IEEE Transactions on Industrial Informatics*, 2022. (Under review)

2. **J. Ye**, E. Stewart, Q. Chen, et al., "A vision-based method for line-side switch rail condition monitoring and inspection," in *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 0, pp. 1–11, 2021.

3. **J. Ye**, E. Stewart, D. Zhang, et al., "Integration of multiple sensors for noncontact rail profile measurement and inspection," in *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

4. **J. Ye**, E. Stewart, D. Zhang, et al., "Method for automatic railway track surface defect classification and evaluation using a laser-based 3D model," in *IET Image Processing*, vol. 14, pp. 2701–2710, 2020.

5. **J. Ye**, E. Stewart, C. Roberts, et al., "Use of a 3D model to improve the performance of laser-based railway track inspection," in *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 233, pp. 337–355, 2019. (Editor's choice award)

1.6 Thesis structure

The diagram in Fig. 1.1 illustrates the structure of the thesis and the progressive relation of each chapter.
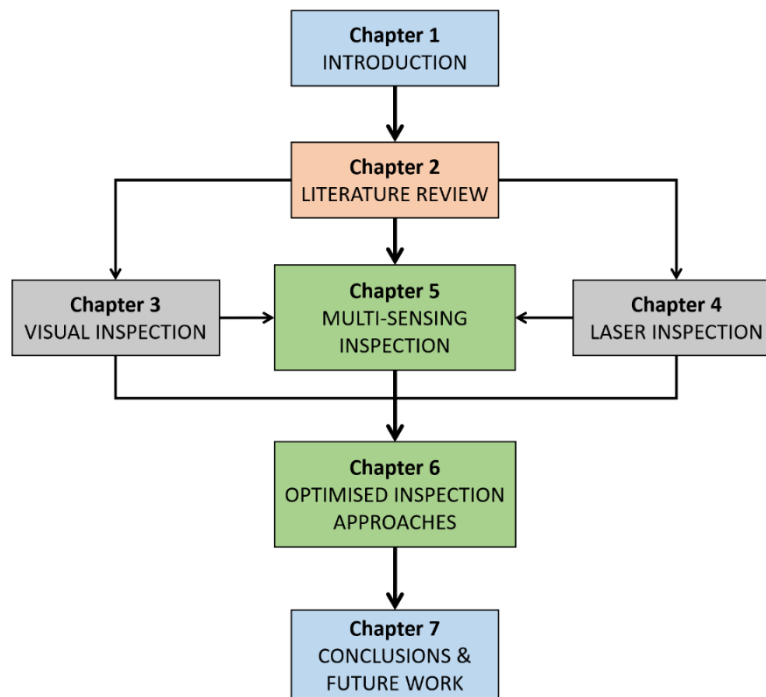
Fig. 1.1 Thesis structure

In Chapter 2, an overview of the requirements for rail inspection in accordance with related standards and existing solutions is given. Related works are reviewed from two perspectives: (1) inspection systems and (2) specified algorithms. From the system perspective, existing vision- and laser-based optical inspection systems are reviewed in detail. From the algorithm perspective, advanced inspection algorithms are introduced. Based on the existing gaps from these two perspectives, multi-sensing technologies are introduced as potential solutions.

In Chapter 3, a line-side rail switch condition monitoring system that was developed by the author for an industrial project is used as a case study, to further explore the characteristics of vision-based methods for rail inspection. The strengths and shortcomings of the system are analysed based on the field tests carried out on the Hefei to Anqing high-speed rail line in China.

In Chapter 4, a prototype that mimics the structure of most of the existing laser-based rail inspection systems, using a 2D profile laser sensor as an onboard measuring unit, is developed as a case study. In this configuration, both 2D measurement and the

feasibility of 3D measurement with a dedicated 3D reconstruction method are discussed to explore the characteristics of laser-based rail inspection.

In Chapter 5, a new prototype for a laser-based multi-sensing system is proposed based on the discussions in Chapters 3 and 4. Details of the system design, the integration of multiple sensors, and a dedicated registration algorithm for complete 2D rail profile measurement are presented. Testing and evaluation results of the new prototype as a hand-held system under the scope of the S-CODE project are discussed.

In Chapter 6, two new strategies for automatic rail inspection are introduced and tested when using the proposed system as an onboard 3D measuring unit. These include end-to-end 3D pixel-level surface defect detection and characterisation and 3D model-based automatic surface defect classification and evaluation.

In Chapter 7, the research work and findings of this research are summarised. The main contributions are concluded. The limitations of this research and potential solutions are also analysed.

2 LITERATURE REVIEW

This chapter presents an overview of the current state of railway track inspection technologies. The specific rail faults of concern in this research are introduced first, based on related standards in the UK and on some European standards as supplements (2.1.1). Subsequently, a brief review of existing inspection methods is given (2.1.2). As the main inspection units and research targets in this research, vision- and laser-based inspection technologies are comprehensively reviewed in two separate subsections (2.2 & 2.3). Existing gaps when they are applied as rail inspection units are analysed, through which multi-sensing technologies are identified as one of the potential solutions to improve their performance. Accordingly, related multi-sensing technologies are introduced in the following subsection (2.4). It should be noted that relevant literature in the area of railways is limited; therefore, similar research in other domains is also reviewed to widen the horizon.

2.1 Overview of railway track inspection processes and requirements

2.1.1   Rail defects and inspection targets

With the development of high-speed rail and the increasing traffic load and density, rails are subjected to increasing stresses. In addition to natural degradation, these stresses can also result in localised rail defects, which endanger the safe operation of trains. Rail defects can be categorised into different types according to their initiation, position, and development. Fig. 2.1 illustrates the basic structure of a rail and some terminologies used in this research, to ease detailed categorisation of defects. From the cross-section view, a rail can be divided into the rail head, rail web, and rail foot. Defects lying above the horizontal plane or along the vertical plane as shown in the diagram are usually classified as a group that is visibly identifiable. These defects normally appear on the rail head surface at the point of interface with the wheels. Other defects that originate from the transverse plane are normally internal, and thus are not visibly identifiable until they develop sufficiently to penetrate the rail surface, which could ultimately lead to a broken rail and requires a rail replacement. With the use of advanced alloys and the development of the manufacturing process in modern railway systems, internal defects have been considerably reduced while, in comparison, surface defects are still difficult to control as they can be derived from the long-term rolling contact between the rail and wheels [5]. Thus, they have always been a major safety concern and a dominant cause of rail failure on all types of railway systems. For

example, a passenger train derailed at Hatfield, UK in 2000, as a result of a broken rail induced by surface defects, which resulted in 4 deaths and 39 injuries; in the US, statistics from the Federal Railroad Administration (FRA) indicate that the safety implication of rail surface defects is responsible for approximately 100 derailments each year [12].
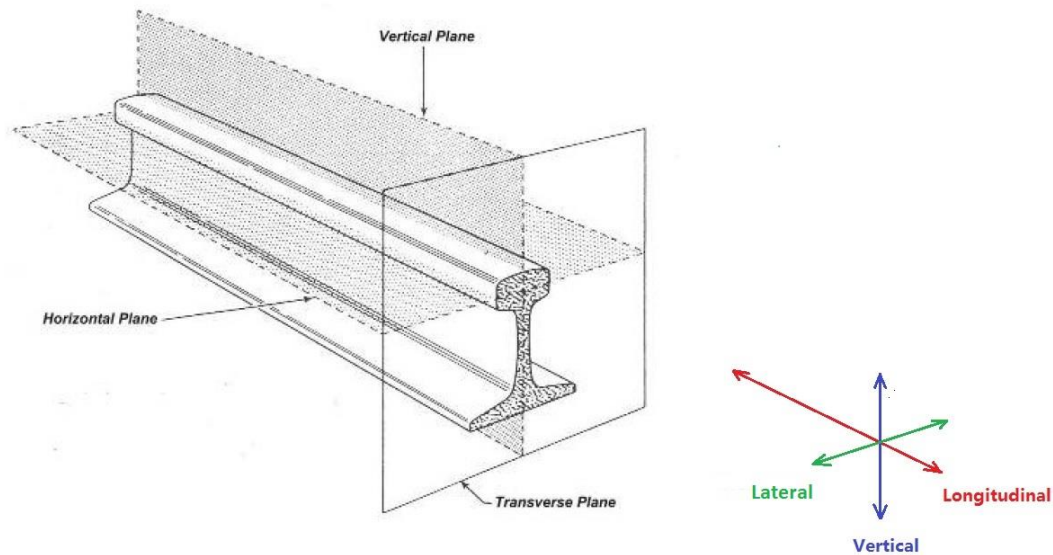
Fig. 2.1 Basic structure of a rail with specific terminology to describe the directions and planes

In addition to plain tracks, surface defects can also be distributed on switches & crossings (S&C). Fig. 2.2 shows the simplest arrangement of a single S&C and its main components. S&C act as nodes within the railway network and enable trains to be guided from one track to another [13]. Their role at junctions within the network and their complicated structure expose S&C to a higher working load than the plain track, and thus they tend to have a shorter life span. Statistics from the International Union of Railways indicate that the maintenance cost of a single S&C unit is equivalent to that for approximately 0.3 km of plain track [14]. Not including the switch machines, the maintenance and replacement costs for the mechanical aspects of S&C account for a large proportion of the ongoing costs [15]. As a transfer area, the discontinuity in geometry such as at the switch rails and the crossing nose (highlighted in red in Fig. 2.2) exposes them to high-impact loads as trains pass through, which can accelerate the initiation and propagation of surface defects. Consequently, surface defects on S&C are also considered in this research.
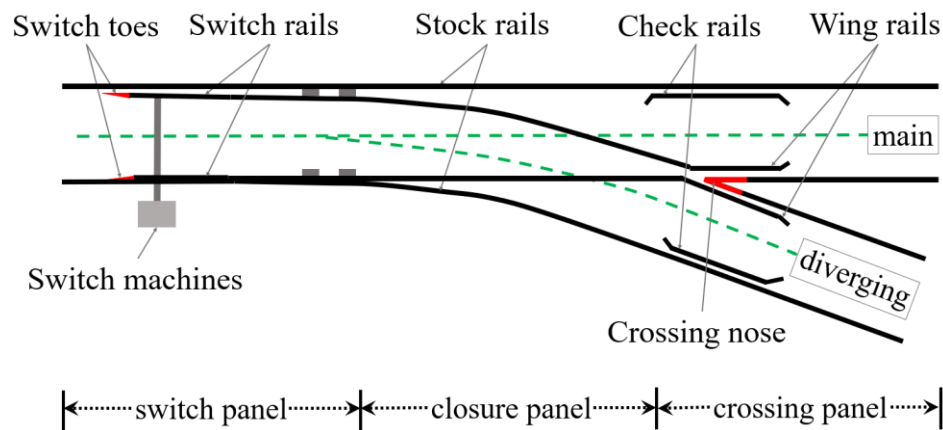
Fig. 2.2 Simplest structure of a single S&C

In recent years, extensive research regarding surface defects has been carried out around the world. This mainly includes the initiation, propagation, inspection, and management of surface defects. Among these tasks, rail surface defect inspection is of particular significance, as it directly ensures the safe operation of rail track systems. The process of rail surface defect inspection mainly involves defect detection, localising detailed regions of defects; and defect classification, determining specific categories of defects. These works support maintenance schedules and ensure that appropriate remedial actions can be applied. Consequently, it is necessary to give a clear view of different types of rail surface defects including their initiation, distribution, and characteristics, which can help to determine appropriate inspection strategies. Table 2.1 summarises the categorisation of rail surface defects used in this research, which include cracks, squats, and shelling. These are also known as rolling contact fatigue (RCF) defects [16]. Abrasion and deformation are also considered as another common type of surface defect. It is worth noting that different rail standards may have different categories or terminologies for rail defects, but they are all aimed at assisting the inspection procedures. Other methods of categorisation can be found in [12, 14, 17].

Table 2.1 Categorisation of surface defects in this research

Cracks
Initiation
Cracks, also known as gauge corner cracking (GCC), are usually caused by stresses and slip force on the rail surface due to the poor wheel–rail contact conditions.
Characteristics
This type of RCF defect is normally characterised by a cluster of small parallel cracks in the gauge corner. Based on the wheel–rail contact condition, the cracks can appear with an angle of 35° to 70° relative to the longitudinal rail axis, against the running direction of the train. Initially, the cracks run a few millimetres into the interior of the rail. Without preventative measures, they propagate quickly and then generate gauge corner collapse. In extreme cases, the cracks can grow transversally and may lead to rail breakage.
Distribution
Generally, cracks are distributed at the gauge of plain tracks and S&C. The pictures below show some examples of GCC of different severity and typical propagation in a cross-sectional view. Cracks in their early stages like (a) can be controlled by preventative measures such as rail grinding. If this is not done, they can deteriorate into a plastic flow and gauge corner collapse as in (b), leading to rail replacement.

(a)   (b)   (c)

Squats
Initiation
Squats are normally shaped from micro-cracks and localised reduction of the strength of the material.
Characteristics
This type of defect is generally characterised as a shallow depression in the centre of the railhead with a circular arc or V-shape. Usually, squats appear with a depth of approximately $1-2\,mm$. When they grow to $3-5\,mm$, they can turn down transversely, eventually leading to a break of the rail.
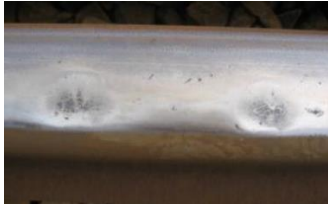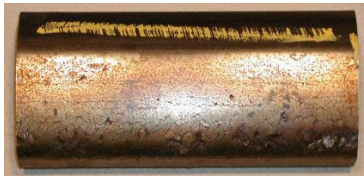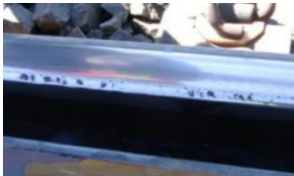
Distribution
Squats can be distributed on the running surface in the centre of the railhead or gauge corner of both plain tracks and S&C. Some typical examples of squats are shown below.

(d)	(e)	(f)

Shelling

Initiation
Shelling refers to the loss of a piece of metal from the rail as the result of either surface cracks linking up with other cracks below the rail, or collapse of the rail gauge corner due to localised loss of structural integrity and poor wheel–rail profile matching.

Characteristics
Rail shelling firstly appears as dark spots due to the merging of surface cracks. It can then quickly propagate to a depth of $2 - 8\ mm$ as shelling or peeling of material, especially on rails in curves. If preventative measures are not taken early, it can continue to grow downward to become transverse defects and finally lead to rail failure.

Distribution
Shelling can be distributed on both plain tracks and S&C, randomly spaced out over the gauge corner. Some examples of shelling in different stages are shown below.

(g)	(h)	(i)

Table 2.1 introduces three common types of rail surface defects. In this research, abrasion and deformation are considered as another type of surface defect, as they are also visually identifiable. A particular form is also known as lipping, as defects often appear as plastically deformed lips. Compared to the other surface defects listed in

Table 2.1, abrasion and deformation have a large impact area. They are often distributed in tight curves and discontinuous connections, where higher impact loads are involved as train wheels pass through. Abrasion and deformation may be initiated differently, but they can both result in partial rail profile changes which lead to abnormal gauge and non-optimal contact with the wheel flange. These endanger the safe operation of passing trains. Fig. 2.3 shows some examples of this type of defect on a switch rail and a plain track with a cross-section view for clarity.

(a)   (b)   (c)
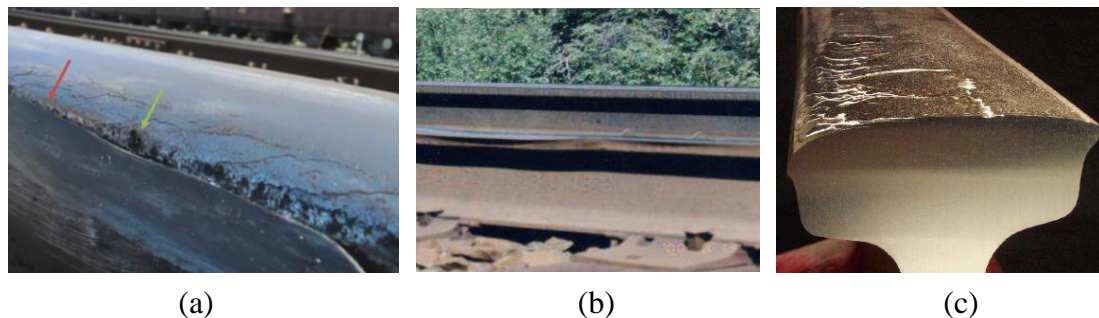
Fig. 2.3 Abrasion and deformation of the rail: (a) a damaged switch rail [9]; (b) and (c) deformation of the underside of the railhead [12]

2.1.2 Existing and state-of-the-art inspection and monitoring solutions

Conventionally, the quality of rails in service is mainly reliant on visual inspections conducted by humans, and sometimes with the use of mechanical contact tools and profile gauges. Human-conducted inspections, however, are normally characterised by low efficiency, subjectivity, and having safety concerns [9] [14] [18]. Therefore, the increasing volume of railway networks and the ever-shortening maintenance window available in modern railway systems have posed challenges to conventional inspection methods. These have triggered the development of automatic and non-destructive techniques (NDT). Compared to manual inspections, NDT solutions are mainly based on advanced sensing technologies. Thus, rail conditions can be sensed more efficiently with less risk of errors. Meanwhile, the digitalisation of sensing data further eases data management and makes it possible to perform automatic data analysis.

According to the inspection targets, e.g., surface defects or internal defects, numerous NDT solutions have been applied to improve the performance of rail inspection. Many review articles about these techniques can be found [19] [20] [21] [22]. In this research, since the inspection targets are rail surface defects, four common techniques applied for surface defect inspection are introduced in detail.

Table 2.2 Common NDT techniques for rail surface defects

NDT technique	Fundamentals	Targets	Systems available
Eddy current	Electromagnetic effects	Surface and near sub-surface	Manual and high-speed (up to 100 km/h)
Ultrasound	Ultrasonic reflections	Surface and sub-surface	Manual and high-speed (up to 70 km/h)
Vision	Passive optical reflections	Surface	Fixed-point and high-speed (up to 200 km/h)
Laser	Active optical reflections	Surface	Manual and high-speed (up to 200 km/h)

As summarised in Table 2.2, the eddy current (EC) technique uses electromagnetic effects to identify surface and sub-surface defects. In addition to rail inspection, it has been widely applied in many other industrial fields, such as oil pipelines and aviation [23]. For rail inspection, EC is normally based on a probe consisting of a primary coil and a sensing coil running along the railhead surface. By feeding an alternating current (AC) to the primary coil, a magnetic field can be formed between the probe and the rail surface. In normal cases, the impedance of the probe remains constant. When the probe passes over a defect, the magnetic field will change and generate an induction current near the surface, which fluctuates corresponding to characteristics of defects, such as the size and the depth. This can change the impedance of the probe, which can be monitored as induced voltage. According to the skin effect of the magnetic field, EC can detect defects in the range from the surface to near sub-surface of around 7 mm (may vary depending on the conductivity of materials and the lift-off distance between the probe and the rail surface). Currently, it has been applied for rail inspection in many forms, such as an EC-based walking stick, an inspection trolley, and onboard systems, to carry out rail inspection manually or at high speeds up to 100 km/h. However, several limitations of EC that can affect its performance in practice have also been demonstrated. The first important limitation is stability. It can be easily affected by the complicated conditions of the rail such as variations of surface conductivity and uncontrollable interference signals. The second limitation is that a stable lift-off

distance (1–2 mm) between the probe and the rail surface is required to ensure accurate inspection, which is unattainable in practice.

Ultrasound is another technique that has been applied for rail inspection. It is mainly based on high-frequency sound waves (from 2.25 to 5.0 MHz) from piezoelectric transducers [24]. By pointing several transducers at different angles, the entire rail can be covered. Rail defects can be detected by identifying abnormal reflections. Compared to EC, ultrasound has better penetration and thus it is better at detecting internal defects. This, in turn, makes it weak at detecting surface defects, especially small ones ($<$ 4 mm). In addition, couplants such as water are typically required between the transducers and the rail surface to assist the transmission of ultrasonic signals. Consequently, ultrasound-based systems have a relatively lower inspection speed. In most applications, ultrasound is often used in combination with EC to increase system capability.

EC and ultrasonic techniques have inherent limitations for detecting surface defects in terms of both efficiency and accuracy. Although there are many variants, such as alternating current field measurement (ACFM), electromagnetic acoustic transducers (EMAT), laser ultrasound, and air-coupled ultrasound, they have not improved enough for mainstream surface defect detection [12].

In recent years, with the rapid development of relevant hardware and constantly improving processors, vision and laser techniques have increasingly been applied in industrial areas. Vision inspection mainly refers to passive imaging techniques represented by high-speed digital (HD) cameras, such as charge-coupled device (CCD), complementary metal-oxide semiconductor (CMOS), and infrared cameras [25]. These sensors contain an array of photosensitive chips which act in the same way as the retina of human eyes and can transform reflected light from targets into electrical impulses. The received electrical signals are eventually converted to digital pixel values. An image of the target can then be formed with this array of pixel values. With the emerging high-resolution and high-sensitivity sensors, high-quality images can be captured in most scenarios. This makes it possible to use HD cameras to take the place of human inspectors and improve inspection efficiency at a reduced cost. Another sensing technique, using active optical reflections, is the laser. For surface defect inspection, laser mainly refers to laser triangulation sensors [26]. By using special lenses, the active light source can be enlarged to form a static laser line, which can then be projected onto the target surface. The diffusely reflected laser light is then received and mapped onto the sensor matrix. Based on the fundamentals of triangular reflection, the distance information and the position along the laser line can be

calculated. This information can then be output digitally to represent the geometry and the surface condition of the target.

Compared to those NDT based on electrical and ultrasonic signals, laser and vision technologies are truly non-contact, with a flexible lift-off distance. This makes them more compatible with deployment at higher inspection speeds. Furthermore, a fundamental of optical reflections determines that they can intuitively reflect the surface conditions of the target, which is closer to visual inspection by humans. Therefore, vision and laser techniques have increasingly been applied for automatic surface defect inspection in recent years. However, laser and vision technologies also have their shortcomings for surface defect inspection. For example, as they are based on optical reflections, they can be affected by some external factors such as surface cleanliness and external lighting conditions in specific application scenarios. To further identify their strengths and existing gaps for surface defect inspection, especially for rail, the following two subsections primarily review their applications in railway systems in detail.

2.2 Vision-based railway track inspection and condition monitoring technologies

Vision-based inspection and condition monitoring systems normally consist of subsystems that effectively integrate data acquisition technology, monitoring technology, and image-processing technology. In recent years, with the development of computer vision technology and the increasing demand for automation, vision-based inspection and condition monitoring technologies have been applied in many areas, such as transportation, industrial manufacturing, and medical imaging [27]. In railway systems, they have also been applied to a wide range of scenarios from passenger flow to railway asset inspection and monitoring [20]. Among these, railway track inspection and condition monitoring play a vital role in ensuring the safe operation of the railway network. Vision-based inspection and condition monitoring systems have the advantages of comparatively high operation speeds, ease of deployment benefiting from the truly non-contact property, and low cost in terms of hardware configuration and the ability to substitute human operators, and thus have received much attention. The following subsections review some representative systems and corresponding inspection strategies.

2.2.1 Existing systems

To carry out railway track inspection and monitoring, the majority of existing systems use HD cameras as an onboard unit, installed under either an inspection train or a trolley.

Meanwhile, considering changes in natural light, external illumination sources are often coupled with the camera to ensure illumination equality and reduce the effect of variation in natural light. Fig. 2.4 shows two representative examples of the onboard vision unit. Specifically, Fig. 2.4 (a) is a real-time visual inspection system developed by Li and Ren [28], which is equipped with a high-speed line-scan camera and four LED lights as lighting sources. The system was designed to be installed under the train and it is claimed to be able to carry out rail surface defect inspection at speeds up to 216 km/h. Fig. 2.4 (b) is an automatic railway visual detection system (RVDS) developed by Zhang et al. [29], which is also equipped with a line-scan camera and a pair of lighting sources. An additional vibration slot is used for vibration damping. It should be noted that although lighting sources are commonly used, such vision-based systems are still inevitably affected by natural light and any contamination on the rail such as rail grease and rust, which have often been claimed as the main challenges and affect their accuracy in practical applications where these factors are unavoidable.
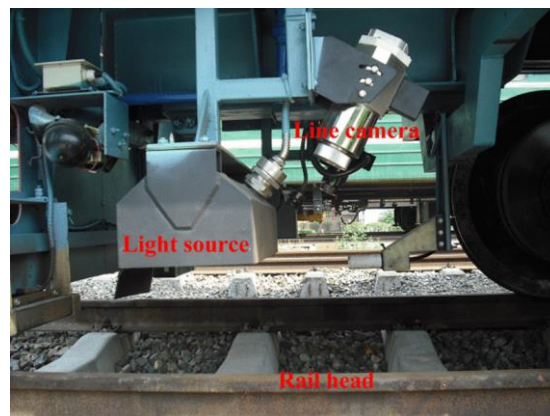
(a)
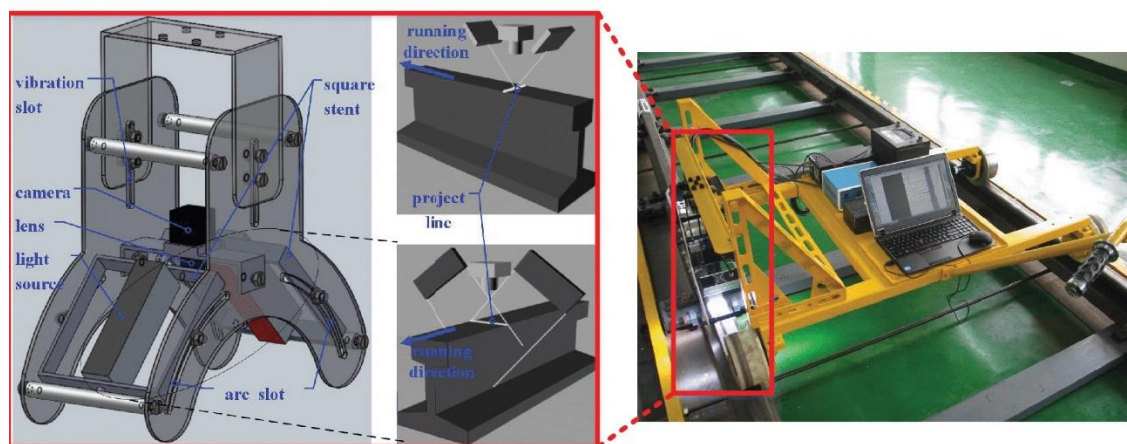
(b)

Fig. 2.4 Examples of vision systems for rail inspection

2.2.2 Advanced inspection methods

The aforementioned vision-based systems are employed for collecting images of the target surface. These are equivalent to the role that human eyes play in conventional manual inspection tasks. To replace or reduce the dependency on humans, these systems also need to be able to analyse and understand useful information from images. The technique that endows vision-based systems with these capabilities is known as computer vision. The founder of computer vision, David Marr, who is a British neurophysiologist, defined that a computer vision task is to understand what is presented in an image and where it is [30]. This points out that the two most fundamental problems for computer vision research are object detection and classification. Therefore, a computer vision-based rail surface defect inspection task can also be divided into two parts: defect detection and defect classification. Specifically, given an input image, defect detection tells where the defect is, and defect classification determines the specific type of defect. In addition to rail surface defect inspection, this rule is universal for surface defect inspection tasks in most different areas, such as steel plate defect inspection, road surface defect inspection, and medical image analysis. Therefore, to widen the horizon, some advanced surface defect detection and classification solutions in other domains are also discussed in this section.

(1) Defect detection

For both vision-based defect detection and classification, feature extraction is a core component. In an image, features refer to useful information that can help to distinguish objects from one another. These could be a distinct colour or a specific shape such as a straight line, circular shape, or edge. In computer vision, feature extraction is the process of finding the features in an image and describing the features with quantitative attributes that a computer vision system can understand [27]. For example, given a colour image, which normally consists of three channels (red, green, and blue – RGB), each pixel (p) of an RGB image can be expressed as a 3D vector as in (2-1).

$$p = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \tag{2-1}$$

where p_1, p_2, and p_3 are the intensity values of the red, green, and blue colours of a pixel.

If the colour of a defect is discriminative, then the region that corresponds to the defect can be represented as a set of feature vectors in 3D space. Similarly, the edge of a defect can normally be detected and described by the gradient of the brightness of neighbouring pixels. Obviously, these simple descriptors cannot cover complex conditions in the real world such as different surface textures and influences from ambient light. To deal with the complex circumstances in practice, more explicit descriptors are required. In recent decades, a lot of feature descriptors have been developed such as histogram properties [31], morphological operations [32] [33], and spatial-frequency analysis [34]. Some of them have also been applied for rail surface defect inspection, for example the discrete rail surface defect detection method proposed by Q. Li and S. Ren [28]. As shown in Fig. 2.5, this first extracts the region of the rail track from the input image based on the fixed width and higher average intensity of the rail track. Subsequently, a local normalisation based on the regional mean and standard deviation is applied to eliminate illumination inequality and enhance the distinction between defective and defect-free regions. Eventually, defects are localised by analysing the spatial intensity distribution. The fundamentals of intensity analysis are that pixels corresponding to the defect are clustered with low intensity, pixels corresponding to defect-free regions are spatially uniform with high intensity, and other pixels that correspond to noise are randomly distributed with random intensity. H. Zhang et al. claimed fast and accurate surface defect detection using a curvature filter and improved Gaussian mixture model (GMM) based on Markov random field
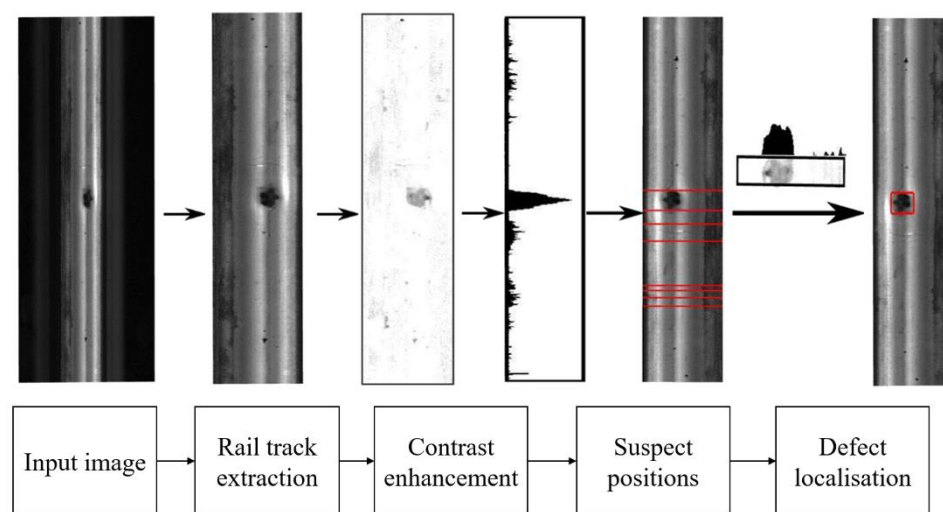
Fig. 2.5 Framework of a rail surface defects detection method based on intensity analysis

(MRF) [29]. It is based on a series of improved filters. Fig. 2.6 illustrates its framework. Firstly, the region of interest (ROI) that contains the rail surface is extracted based on greyscale histogram analysis. Secondly, a grey equalisation filter is used to smooth the non-uniform brightness of the image caused by illumination inequality and orbital reflectance inequality. A curvature filter is then applied to remove noisy points while retaining the details of defects. Finally, the filtered image is fed into an improved Gaussian mixture filter based on the MRF to incorporate the spatial correlation among pixels for defects segmentation.
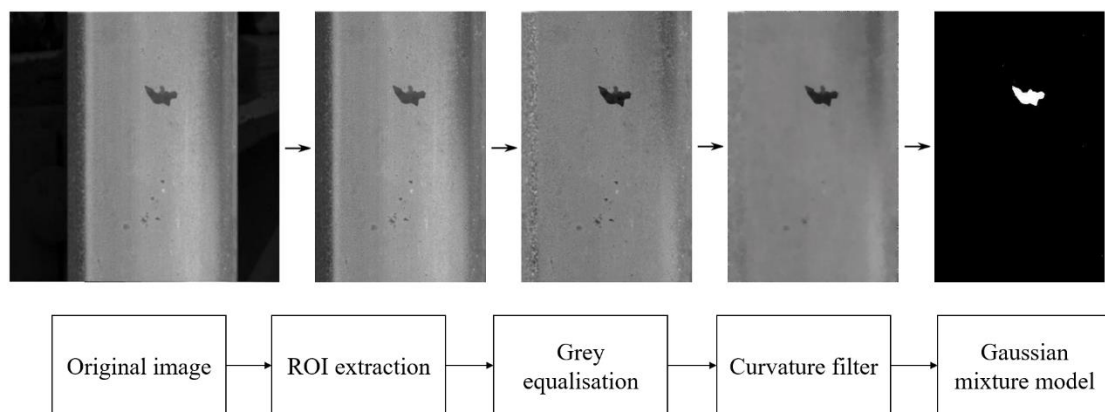
Fig. 2.6 Framework of a rail surface defect detection method based on curvature filter and improved Gaussian mixture model

Such methods applied for rail surface defect detection are limited. Therefore, those representative methods that have been applied for other surface types were also reviewed. To give a better understanding and to help navigate through different methods, the categorisation of these feature descriptors is necessary [35]. According to the nature of the methods, this research summarised several review articles and categorised existing surface defect detection methods into three groups: statistical, spectral, and model-based. Fig. 2.7 lists the representative methods of each category. It is worth noting that numerous methods are not listed and the categorisation in Fig. 2.7 is intended to be illustrative but not comprehensive. Some research may combine methods in different categories, e.g., [33] combines a Gabor filter and morphological features for detecting pinholes in the steel manufacturing industry, and thus taxonomy can vary in different researches. Several other comparative review articles can also be referred to for details [20] [36] [37].

```
                    ┌─────────────────────────────┐
                    │   Surface Defect Detection  │
                    └─────────────────────────────┘
```

Statistical	Spectral	Model-based
Histogram properties	Fourier transform	Markov random field model
Co-occurrence matrix	Gabor filters	Fractal model
Local binary pattern	Wavelet transform	Weibull model
Grey-level statistics	Multiscale geometric	
Thresholding	Hough transform	Other latest reported models
Morphological		
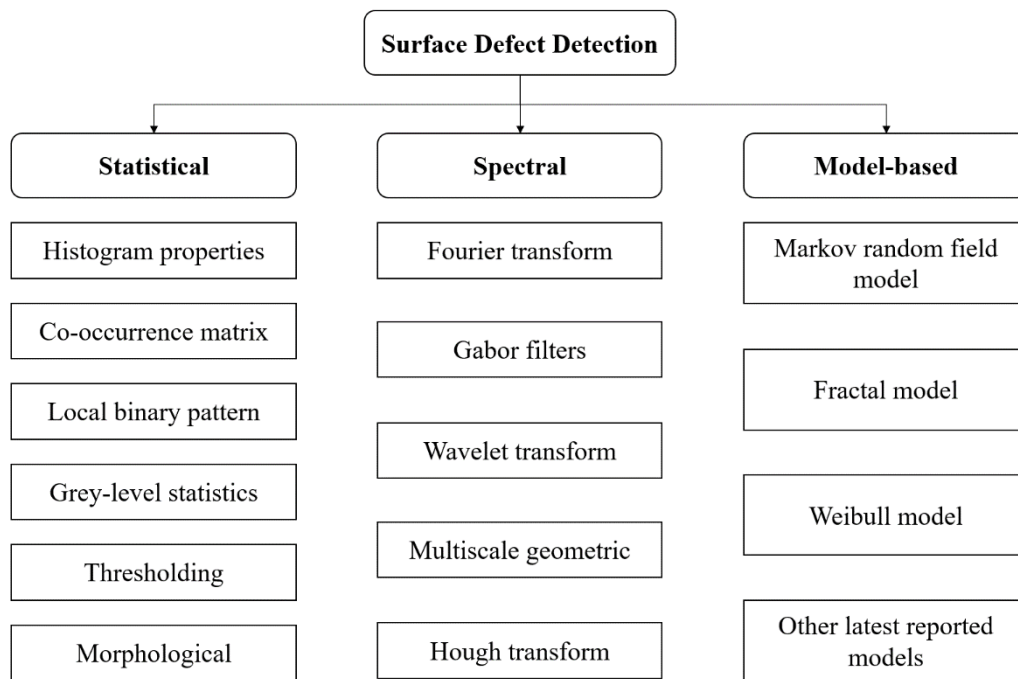
Fig. 2.7 Categorisation of existing vision-based surface defect detection methods

Among these methods, the fundamentals of statistical methods as used in vision processing are to measure the spatial distribution of pixel intensities directly. The strengths of statistical methods are that they are low cost and easy to understand and implement. Therefore, they have been frequently used for various tasks such as steel plates, hot-rolled strips, and wood products [38] [39] [40]. However, the weaknesses of statistical methods are also obvious. They are normally sensitive to noise and easily disturbed by defects and regions with subtle intensity transitions. Spectral methods are applied in the frequency domain, where the Fourier spectrum represents an image by obtaining global features in transform domains (i.e., spatial domain and frequency domain). Since the methods are in transform domains, they are less sensitive to noise and intensity variations than statistical methods in the spatial domain. However, it is often reported that spectral methods lack local information, and thus they may have difficulty in detecting defects with irregular shapes and relatively small sizes [33] [39] [41]. Apart from statistical and spectral methods, there are also a group of methods that represent the original texture distribution of image blocks in low-dimensional space by using texture models, known as model-based methods. These methods are normally claimed to perform better for the detection of diverse defects. However, building noise-robust, explainable, and low computational complexity models is difficult [42] [43].

(2) Defect classification

Defect classification is another essential part of rail surface inspection. It aims to assign defects to a series of known classes as described in standards or specific maintenance documents [44]. In computer vision, this is normally based on the pattern that is formed by a set of features of defects that share common properties, and thus it is also known as pattern recognition or pattern classification. Since the features that help to achieve defect detection can potentially provide effective information for defect classification, defect classification is normally a further step after the defect is detected. There are two common approaches for defect classification: (1) classification based on matching patterns with specified rules and (2) classification based on statistical classifiers [36]. The first type is suitable for some simple classification tasks whose features are unique and low-dimensional. In this case, it is possible to design a set of rules for different classes and classify the defect by matching the corresponding pattern against the rules. For example, as shown in Fig. 2.8, Z. Xiong et al. built a decision tree that consists of five features and specified rules for classifying five types of rail surface defects [45]. Features on a route that navigates to a class form the corresponding pattern for classification. However, practical classification tasks are normally much more complex, such as those with linear non-separable data, for which more features and more explicit rules are required. In this case, it is intractable to design specified rules manually. In the second category, classification is based on finding the mathematical regularity
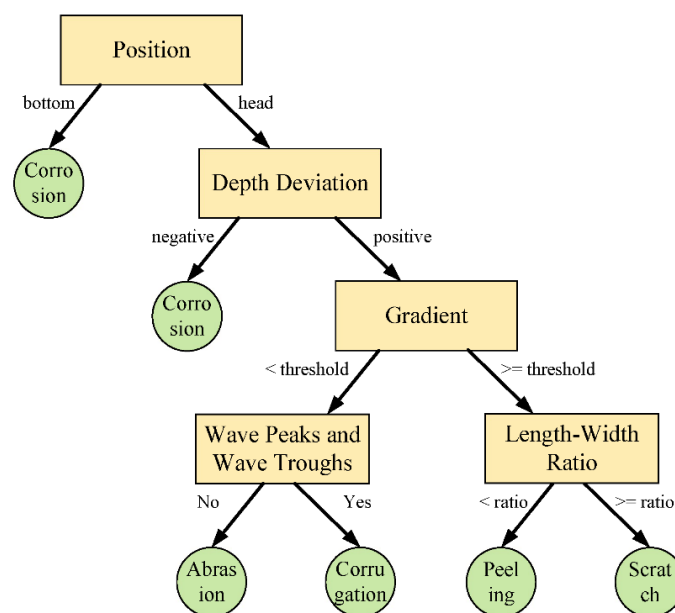
Fig. 2.8 A simple decision tree-based rail surface defect classification method

between features vectors (x) and pre-defined classes (y) of the data. The function $y' = f(x)$ that defines the regularity is called the classifier [36]. This is also a form of machine learning in which the classifier is trained on a group of training datasets based on their spatial distribution properties. In other words, the classification process is data-driven rather than explicitly programmed. The trained classifier can then be applied to classify an unknown data set, and thus it is more efficient for the classification of a relatively complex and large data set. From Bayesian to support vector machine (SVM), a large number of classification models have been developed and applied for various surface defect classification tasks. Taking SVM as an example, it defines hyperplanes as in (2-2):

$$f(x) = w^T \emptyset(x) + b \qquad (2\text{-}2)$$

where w^T is the normal vector of the hyperplane, b is the translation determining the distance between the hyperplane and the origin, and $\emptyset(x)$ represents the projected feature vector.

For linearly separable data, SVM classifiers find the hyperplane easily with the introduction of a soft margin. Additionally, linearly non-separable data can be converted to be linearly separable data by projecting them into a higher dimensional space with the use of proper kernel functions. For example, Xue-Wu et al. defined a group of spectral texture measures for classifying reflective metal surface defects. A radial basis function kernel SVM-based classifier was trained with a classification accuracy of up to 91.3% [46].

(3) Deep learning algorithms

Despite the feature processing and classification methods described having achieved good performance in understanding images and extracting useful information for defect detection, they have a common problem: the determination of features relies heavily on human experience, and thus is subjective. For simple and homogeneous defect detection problems, it is possible to handcraft a sufficient number of features. However, it is time-consuming and intractable to handcraft refined features for more complex problems that could be encountered in practice, for example, inhomogeneous background components that have low contrast against defects and surface contaminations that have a similar texture to real defects, etc.

Recent developments in deep learning can overcome these drawbacks and have proven to be effective in many vision tasks [47]. A strong proof would be the annual ImageNet competition, which aims to classify high-resolution images into 1000 categories after training on 1.4 million images. Before 2012, the best score based on classical approaches was only 74.3%. Since the entry of GPU-trained deep neural networks in 2012, the competition has been dominated by deep learning models. By 2015, the best score from deep convolutional neural networks (CNNs) has reached 96.4%, which turns the ImageNet classification task into a solved problem [48]. Different from conventional feature descriptors, deep learning models extract representations through a deep network. This helps to map input data to the expected output targets directly. Fig. 2.9 illustrates a common framework, which most of the popular deep learning models follow. 'Deep' in deep learning refers to the depth of successive layers through which increasingly meaningful representations can be learned from input data (X). The representations of what a layer can extract are determined by the weights of each layer. Initially, random weights are assigned to the network. The network randomly transforms the input (X) to the output (Y'), which is anticipated to be far from the true targets (Y). The foundation of the learning process is achieved by a loss function that computes a loss score between Y' and Y. This allows the optimiser to use the score as a feedback signal, and to adjust the value of the weights a little in a correct direction that will lower the bias between Y' and Y. Such a process, which is also known as backpropagation, can iterate in a training loop, yielding a trained network whose
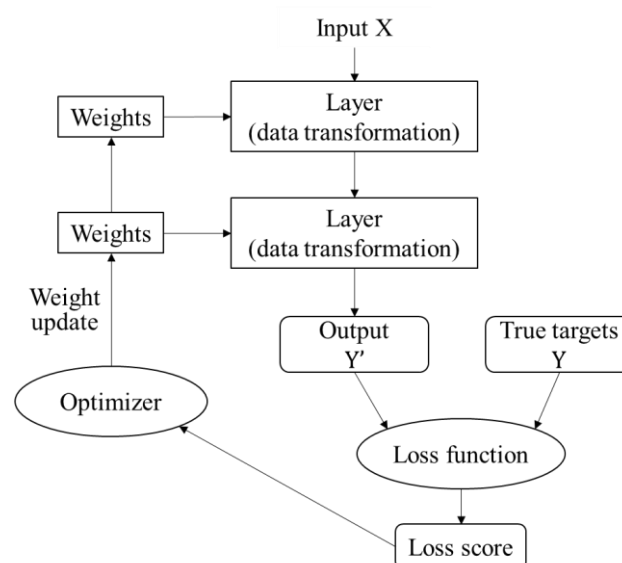
Fig. 2.9 A common framework of deep learning models

outputs are as close as they can be to the true targets with a minimal loss. With a sufficient number of input data (as training data), the trained model will have fine generalisation which can perform well on the new dataset (i.e., validation data and test data).

In computer vision, CNNs, also known as convnets, are universal solutions for vision tasks and have also been widely applied for vision-based surface defect inspection. In a convnet, the fundamental components are convolution layers. Convolutions operate by sliding windows with a specified height, width, and depth over the input sample. Each convolutional operation can transform the corresponding region in the input sample into a feature vector (via the learned weight matrix), which can ultimately form a feature map. With the increase of successive layers, these feature maps can have multiple levels of abstraction with strong representation ability. The handcrafted features from conventional methods, by contrast, are normally the artificial combination of low-level features [47]. Therefore, convnets can perform much better than conventional methods in many applications.

This research categorises the existing deep learning-based surface defect inspection into three groups, namely image-level defect classification, region-level defect detection, and pixel-level defect detection.

(1) *Image-level defect classification* mainly refers to those methods that carry out defect classification at the image level. This kind of model normally has a CNN baseline in parallel with a classifier such as an SVM classifier or a Dense layer with activation functions (e.g., sigmoid and softmax), and classifies input images to corresponding classes (Fig. 2.10). For example, Natarajan et al. [49] proposed a majority voting mechanism (MVM) for metal surface inspection, which fuses features from the last three layers of a deep convnet. These features are then fed into an SVM classifier to classify images of the metal surface into six classes. Liu et al. [50] used a pre-trained convnet in AlexNet architecture for feature extraction and classified RCF defects on the rail surface into three classes. However, the main shortcoming of these methods is that they cannot provide the exact location of defects. Moreover, when there are multiclass defects in one image, the classification accuracy will be affected.
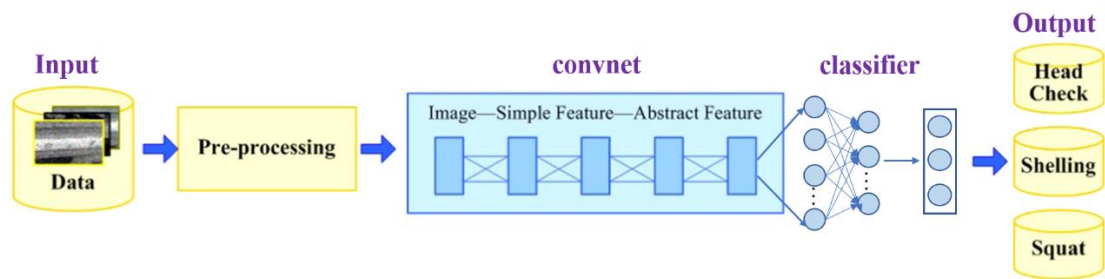
Fig. 2.10 Image-level defect classification (adapted from [50])

(2) *Region-level defect detection* aims to classify and locate certain defects inside the image with bounding boxes. There are mainly two popular solutions: 1) region-based methods and (2) regression methods. The most famous region-based method is the region proposal network (RPN), which maps multiple scale and aspect ratio anchor boxes to the input images. The feature map that comes out from each region has a score indicating whether the corresponding region or the anchor box contains a defect (Fig. 2.11). For example, He et al. [51] proposed a feature fusion network, which combines multi-level features extracted from a convnet in ResNet architecture. The multi-level feature map is then fed into an RPN network to classify and localise steel surface defects. Regression methods are mainly based on bounding boxes that can regress directly to the ground-truth boxes such as you only look once (YOLO) series [52]. The main shortcoming of these methods when they are applied for defect inspection is that they can only provide a coarse bounding of the defect but cannot detect the precise boundary. This, however, is normally required for further defect assessment.
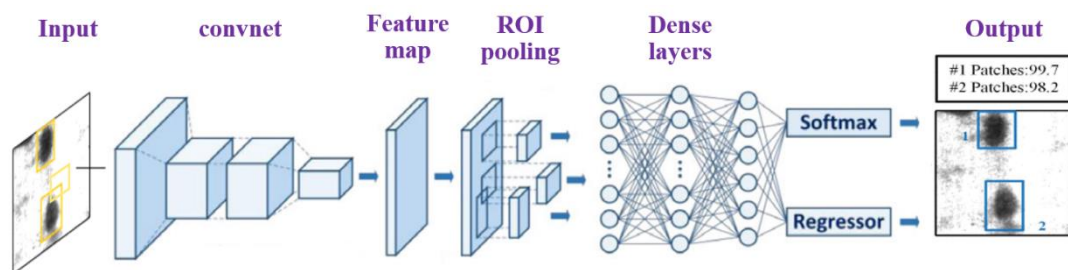
Fig. 2.11 Region-level defect detection (adapted from [51])

(3) *Pixel-level defect detection* is currently one of the most effective defect detection methods. With the improvement of image quality and the computational capacity of hardware, it is possible to classify each pixel in an image to its corresponding

class. Therefore, the region and the boundary corresponding to the defect can be precisely detected with these methods. Currently, the most effective pixel-level detection networks are based on a fully convolutional network (FCN), which normally consists of an encoder and a decoder [53]. The encoder is normally a stack of convolutional layers for feature extraction, while the decoder contains symmetrical de-convolutional layers for up-sampling corresponding feature maps to a segmentation map that indicates the class of each pixel (Fig. 2.12). For example, Yang et al. [54] proposed a convnet that integrates multi-level context information in a feature pyramid way for pixel-level pavement crack detection. Dong et al. [55] used a convnet in VGG-16 architecture for feature extraction through transfer learning, and then fed the fused feature map into a de-convolutional block containing a global context attention (GCA) module and a boundary refinement (BR) module for pixel-level defect detection. Compared with image-level and regional-level methods, these pixel-level methods can achieve both defect localisation and boundary description. However, one shortcoming of these methods is that since the detection is at the pixel level, they are more sensitive to noise. Therefore, it is necessary to have a sufficient number of high-quality images as a training set to ensure the best possible performance of the model for precise pixel-level defect detection; however, this is difficult to achieve for most industrial applications.
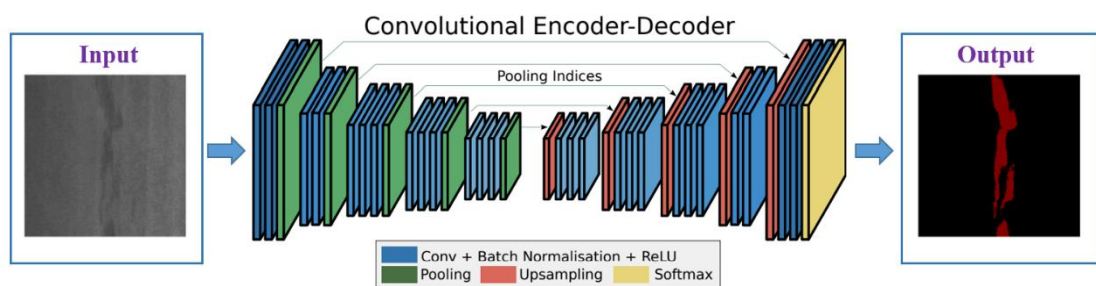
Fig. 2.12 Pixel-level defect detection (adapted from [56])

2.2.3   Limitations of vision-based systems and research opportunities

Based on the discussion above in subsections 2.2.1 and 2.2.2, the main characteristics of vision-based approaches can be summarised. From the system perspective, vision-based inspection systems are normally easy to deploy and cost-effective. Meanwhile, high-resolution and colourful images are becoming achievable in more and more applications with the emerging of advanced optical sensors. From the inspection

method perspective, the strengths of vision-based methods are more prominent. With a wide range of applications, numerous inspection methods or algorithms have been developed. From conventional statistical methods to state-of-the-art deep learning networks, plenty of them have been proved to be effective for applications in many areas such as medical diagnosis, quality control of manufacturing lines, and architectural restoration of cultural relics, etc.

In recent years (typically 2010 onwards), the emergence of big data and constantly improving processors have made deep learning applicable to more and more industrial areas. Surface defect inspection is among these areas. Representative surface defect datasets such as the NEU-Seg dataset [57], MT defect dataset [58], and road defect dataset [54] have produced many promising solutions. However, there are rarely mature solutions for inspecting rail surface defects. The main challenges can be summarised into two points, which also indicate the existing gaps and research opportunities.

1) **Data quality:** The performance of vision-based surface defect inspection approaches normally depends on the image quality. High-quality images are those that provide a clear representation of the surface conditions of the object and are less affected by noise. However, the rail surface in the real world can be dark and oxidised, which result in low contrast and interference between defects and backgrounds (e.g., normal surface in the image). Furthermore, the passive imaging property of cameras determines that high-quality images require stable reflection properties of the rail surface and equality in external illumination. However, the running surface of the rail where defects are mostly distributed can be reflective and present different conditions in different weathers. These hostile factors can lead to the imaging process being distorted, which further reduces the image quality.

2) **Data availability:** To obtain a generic detector for surface defect inspection, a sufficient number of defective samples are necessary, from which relevant features or representations can be built. This demand becomes more insistent with the popularity of deep learning models, which normally need to be trained on a large training set. However, surface defects are randomly distributed on a large rail network. Meanwhile, in modern railway systems, the rail in use is mostly operated in a good condition with a rapid maintenance response. These facts make it intractable to collect a sufficient number of defective samples.

The strengths of vision-based approaches determine their importance for related applications. The existing gaps in their application for rail surface defect inspection also provide research opportunities and a possible route to exploring the application of vision-based approaches in more challenging areas. For example, concerning image quality, robust imaging instruments that have more tolerance to surface contaminations and external lighting variations should be expected to improve the quality of surface characterisation under complicated environmental conditions.

2.3 Laser-based railway track inspection technologies

Laser sensors are among the most popular solution in the industry for applications such as inspection and quality control, due to their robustness, high precision, and truly non-contact properties. In railway systems, 2D laser triangulation sensors, also known as 2D line-structured light, are representative and have been applied for rail profile measurement and inspection. As an illustration, Fig. 2.13 shows such a laser triangulation sensor [59]. For profile measurement, a static laser line is formed using special lenses and projected onto the target surface. The diffusely reflected laser lights are then received by the receiver and mapped onto the sensor matrix, from which the distance information (z-axis) relative to the sensor matrix and the position (y-axis) of each measuring point along the laser line can be calculated by (2-3) [60].

$$\rho \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = A(R,t) \begin{pmatrix} y \\ z \\ 1 \end{pmatrix} \tag{2-3}$$

where (y, z) represents the coordinate of points on the light plane, (u, v) is the coordinate of corresponding points on the sensor matrix, ρ is the scale factor, and A is the intrinsic parameter matrix of the receiver consisting of the rotation matrix (R) and translation vector (t).

These data can then be output digitally to a 2D profile of the target surface, which can be used as a scientific reference for inspection and maintenance works [10] [61]. The following subsections review the applications of laser-based methods for railway track inspection and corresponding inspection strategies.
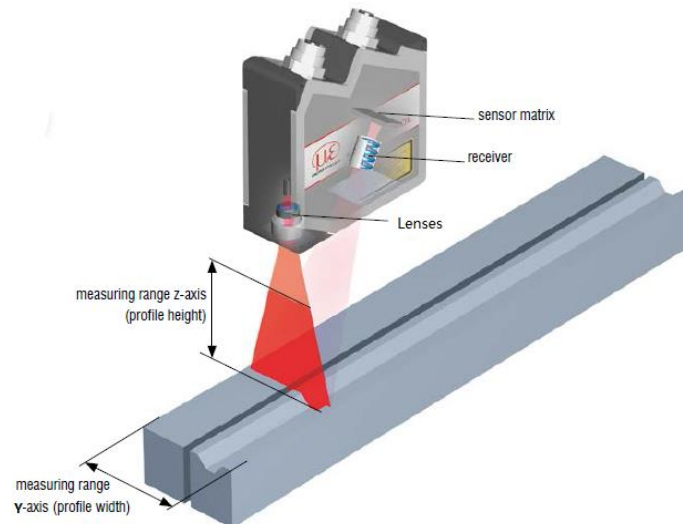
Fig. 2.13 Illustration of laser triangulation sensors

2.3.1   Existing systems

For railway track inspection, laser triangulation sensors are normally deployed in three different ways: i) on board, (ii) on a cart or trolley, or (iii) at a fixed point. The majority of existing systems in the industry use laser triangulation sensors as onboard measuring units to provide digital profile information for the rail, for example Network Rail's New Measurement (NMT) Train in the UK and the High-Speed Comprehensive Train in China [60]. Fig. 2.14 shows an example of the onboard system developed by MERMEC and the measured rail profile [6]. To increase the coverage of the rail profile, other systems may consist of coupled laser triangulation sensors pointing at both sides of the rail. The non-contact and high profiling frequency properties of the laser sensor enable such systems to carry out rail profile measurements at train speeds.

(a)   (b)
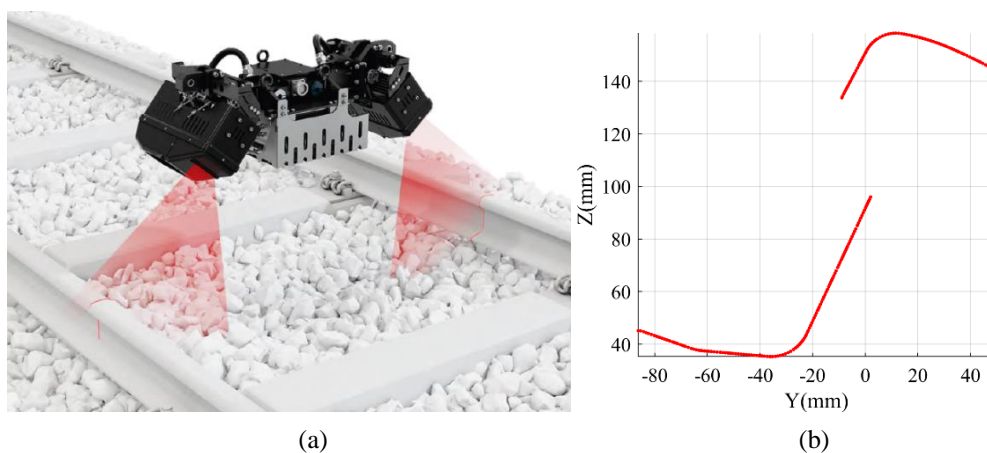
Fig. 2.14 Laser-based onboard rail measurement: (a) MERMEC rail profile measuring system; (b) example of measured rail profile

Alternatively, there are also systems with laser triangulation sensors mounted on a cart or a trolley that can be moved along the rail, for example the mobile robot developed by Loccioni, also known as Felix, for measuring railway S&C dimensions [62]. Xiong et al. presented a laser profiling system with a pair of laser triangulation sensors mounted on the side of a cart to carry out rail profile measurement (Fig. 2.15) [45]. Using a similar framework, Yang et al. [63] installed one laser triangulation sensor to measure rail profiles for rail wear inspection. Compared to train-borne systems, these systems are normally actuated by onboard motors or additional pushing force, and thus the operating speed of such mobile platforms is slow but allows detailed inspection of specific regions.
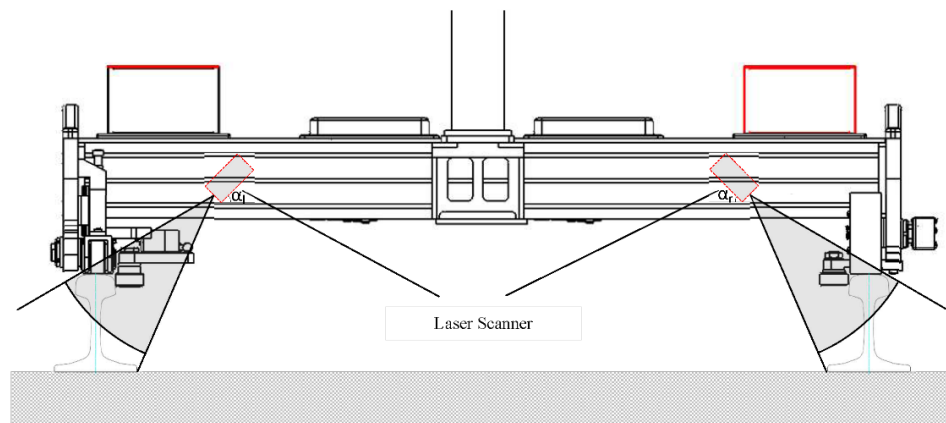
Fig. 2.15 Laser-based rail profiling cart

The two deployments introduced above are mostly applied for inspecting the rail in use. There are also fixed-point systems for quality control of the rail in the manufacturing process. Such systems are normally required to cover the full rail profile, and thus multiple laser sensors are used with each one pointing at a different area of the rail. Fig. 2.16 shows the geometry of a rail profile measurement system (PMS) developed by Molleda et al., which uses four coupled laser projectors and receivers to inspect the quality of rail rolling mills [64].
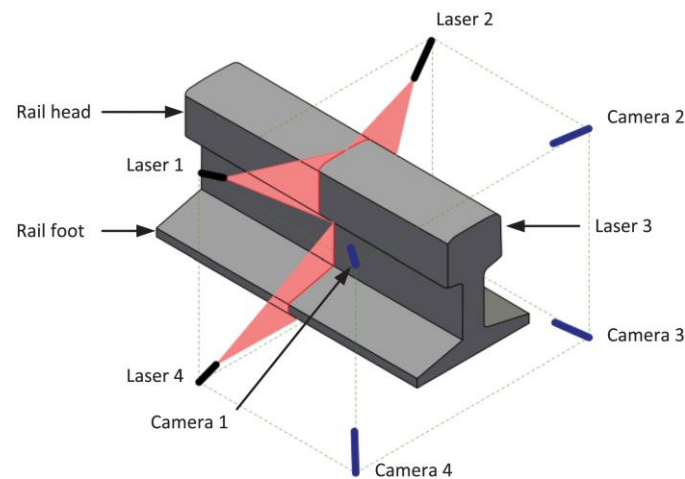
Fig. 2.16 Laser-based fixed-point rail profiling system

It is also worth noting that there are also 3D scanners that project structured light patterns onto the surface of objects, and use a camera array (e.g., binocular vision) to receive reflections so that the 3D information can be obtained. Such systems are normally used for quality inspection or reverse engineering of key components in areas such as the automotive and aircraft industries [65] [66]. There have also been attempts to apply 3D scanners for rail measurement and inspection. For example, Fig. 2.17 shows the 3D scanning trolley developed by Chen et al., which integrates such a scanner to obtain 3D point clouds of the rail surface [67]. Compared to 2D profiles, 3D data are anticipated to be more informatic. However, one shortcoming of such systems is that reference point markers normally need to be attached to the object to align scans for accurate measurement results. Moreover, it inevitably takes more time to generate 3D data, and thus the measurement speed is limited. These affect the practicability of 3D systems for carrying out measurement and inspection tasks for large objects like rails.
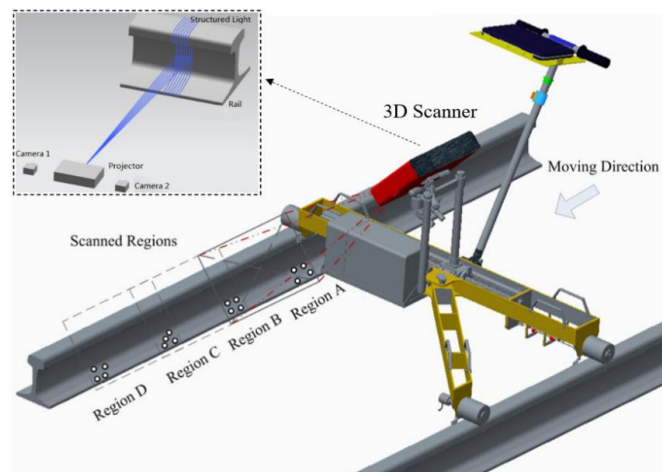
Fig. 2.17 3D laser-based rail inspection trolley

2.3.2   Advanced inspection methods

According to the discussion in section 2.3.1, the majority of laser-based systems carry out rail inspection in the form of 2D rail profile measurement. These 2D rail profiles reflect the rail section geometry information, and thus the most common inspection and wear assessment method is to compare the measured profile with the corresponding standard profile. Since 2D profiles consist of separated data points in the 2D space, the principal process of the comparison is to find matched points between the measured and the standard profiles, and then calculate the transformation matrix between the matched points to align profiles. This is also known as point registration [68]. For rail profile alignment, there are commonly two registration methods used: (i) double circle segment (DCS) [69] and (ii) iterative closest point (ICP) [70]. Fig. 2.18 illustrates the basic principles of DCS, which uses two tangential circles to localise the curved part of the rail web to align profiles. The transformation matrix (T) can then be defined by (2-4), containing the rotation matrix ($R(\theta)$) and the translation offset (T_y, T_z).
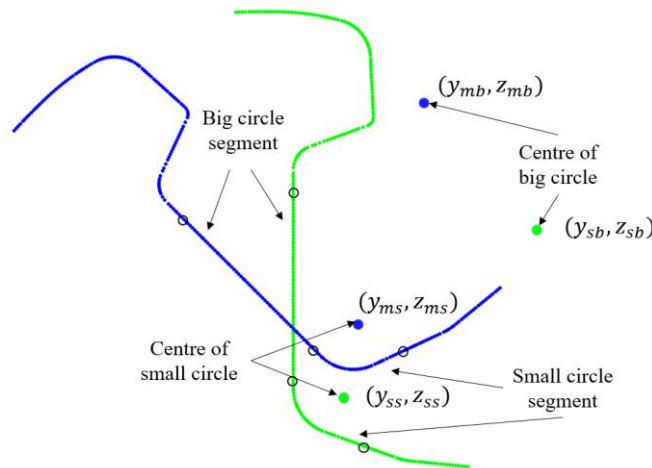
Fig. 2.18 Basic principles of DCS for rail profile alignment

$$T = \begin{bmatrix} cos\,\theta & -sin\,\theta & T_y \\ sin\,\theta & cos\,\theta & T_z \\ 0 & 0 & 1 \end{bmatrix}$$

$$\theta = tan^{-1}\left(\frac{z_{sb} - z_{ss}}{y_{sb} - y_{ss}}\right) - tan^{-1}\left(\frac{z_{mb} - z_{ms}}{y_{mb} - y_{ms}}\right)$$

$$[T_y, T_z] = [y_{ss}, z_{ss}] - [y_{ms}, z_{ms}]$$

(2-4)

Similarly, ICP identifies matched points between profiles using a nearest neighbour search such as a k-dimensional tree (k-d tree) [71], and then uses the least-squares method as in (2-5) as a metric to update the transformation matrix (T) to minimise the squared Euclidean distance ($E(T)$) between profiles for alignment.

$$E(T) = min \sum ||s - Tm||^2 \tag{2-5}$$

where s represents the standard profile, and m represents the measured profile.

There are also variants based on the deformation of classic methods aiming to improve the performance of registration, such as scaling iterative closest point (SICP), reweight-scaling iterative closest point (RSICP), and fast global registration (FGR). For example, Yang et al. [72] compared the performance of ICP, SICP, and RSICP for inspecting rail profile distortion (Fig. 2.19) and compared more registration methods using a sequence of profiles for rail wear inspection (Fig. 2.20) [63].
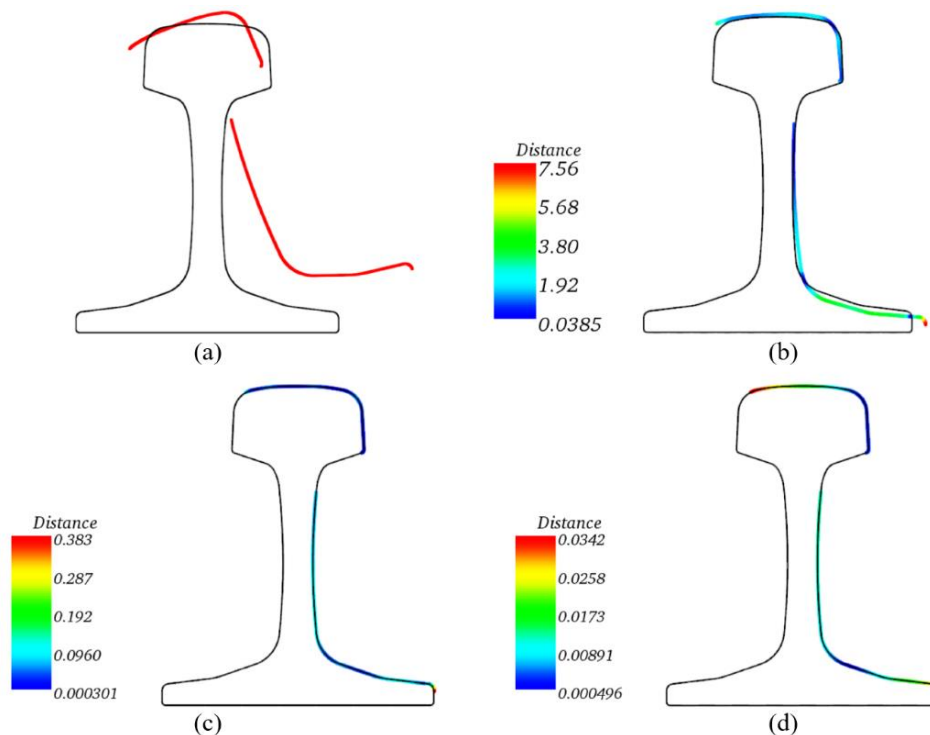
Fig. 2.19 Precision comparison between ICP, SICP, and RSCIP:
(a) original model; (b) ICP; (c) SICP; (d) RSICP
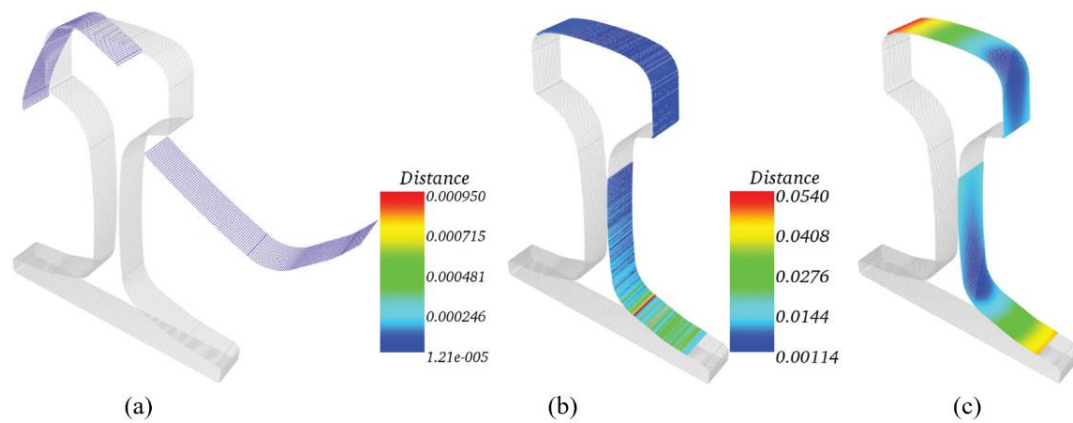
Fig. 2.20 Precision comparison between revised FGR (RFGR) and FGR: (a) original model; (b) RFGR; (c) FGR

As discussed in section 2.3.1, there has also been research using 3D scanners for rail inspection. 3D data points from such systems can also be aligned to the corresponding CAD model for wear inspection. For example, Chen et al. [67] applied a two-step registration (coarse registration and fine registration) for aligning measured 3D rail surfaces to the CAD model, and detected rail wear by visualising the misalignment between corresponding points (Fig. 2.21).
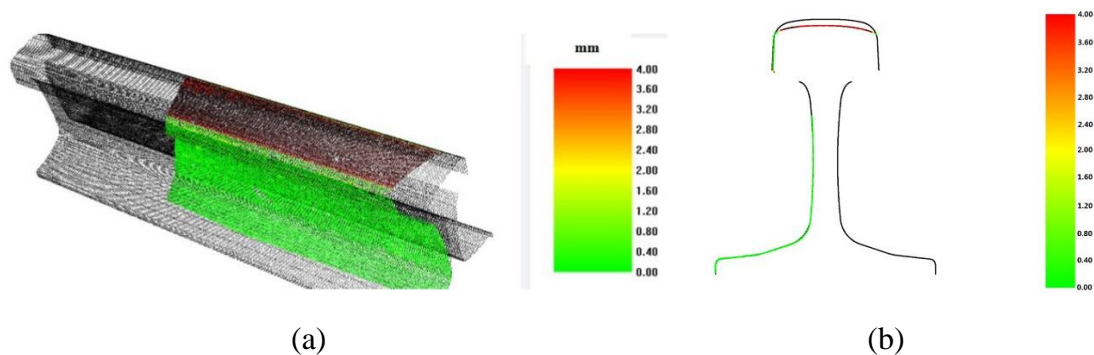
Fig. 2.21 3D comparison with the CAD model and wear assessment

2.3.3   Limitations of laser-based systems and research opportunities

As discussed in the previous subsections (2.3.1 & 2.3.2), laser triangulation sensors are among the most popular solutions for surface geometry measurement and inspection due to their robustness, high precision, and truly non-contact properties. For rail inspection, laser sensors measure the surface geometry of the rail through active optical reflections, and thus they are more stable compared to vision-based systems and rarely affected by surface conditions such as a regional shiny or oxidised surface that a rail can present in the real world. Their application as rail profiling systems has

considerably improved the accuracy and efficiency of rail inspection. However, most laser sensors are deployed as onboard measuring units and the fixture of the laser sensor is reliant on an additional mechanical frame. From a system perspective, such deployments mainly have two shortcomings:

1) **Size:** The existing systems are bulky due to their dependence on additional mechanical frames. This decreases the degrees of freedom of the laser sensor and limits the capacity of the system to carry out more complex measurement tasks.

2) **Integrity:** The integrity of rail profile measurement is difficult to achieve. As an illustration, Fig. 2.22 (a) shows a schematic of existing laser sensor-based rail inspection systems. To measure the rail from a fixed point, there are always 'blind spots' that cannot be seen by the laser sensor. This is also known as the shadowing effect of the laser line, which refers to the fact that the fan-shaped form of the laser line can disappear partially or completely behind steep edges in the targets. This leads to a gap or missing area in the measured profile as shown in Fig. 2.22 (b). This gap is normally ignored in most existing systems as the underside of the railhead is considered to be defect-free. However, in some cases when the discrepancy between the actual profile and the standard profile occurs due to long-term wheel–rail interaction or wear and tear, accurate measurement of unworn areas such as the transition area between the railhead and the rail web is significant for accurate profile registration and wear measurement.

(a)   (b)
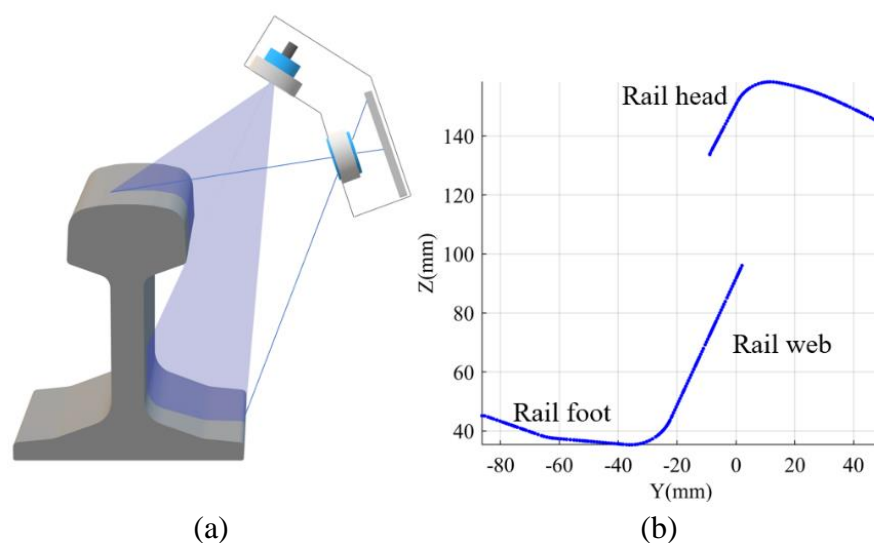
Fig. 2.22 Non-contact rail profile measurement based on laser triangulation sensors:
(a) system schematic; (b) example of measured rail profile

Currently, there are mainly two solutions for this issue: i) filling the gap between the railhead and the rail in post-processing, with a straight line [69] or by dividing the profile from the gap into two continuous parts [73], as shown in Fig. 2.23; ii) using multiple laser sensors: as shown in Fig. 2.16, the missing areas can be supplemented by using multiple sensors with each sensor pointed at a different area of the rail [64] [74]. The former, however, produces an approximation of the missing area, which does not solve the problem in essence. The latter, using multiple sensors, does solve the profile integrity issue. However, the deployment of multiple sensors further increases the system's size, structural complexity, and cost. Therefore, such deployment is normally used for inspecting the quality of rail rolling mills in the manufacturing process rather than the rail in use.
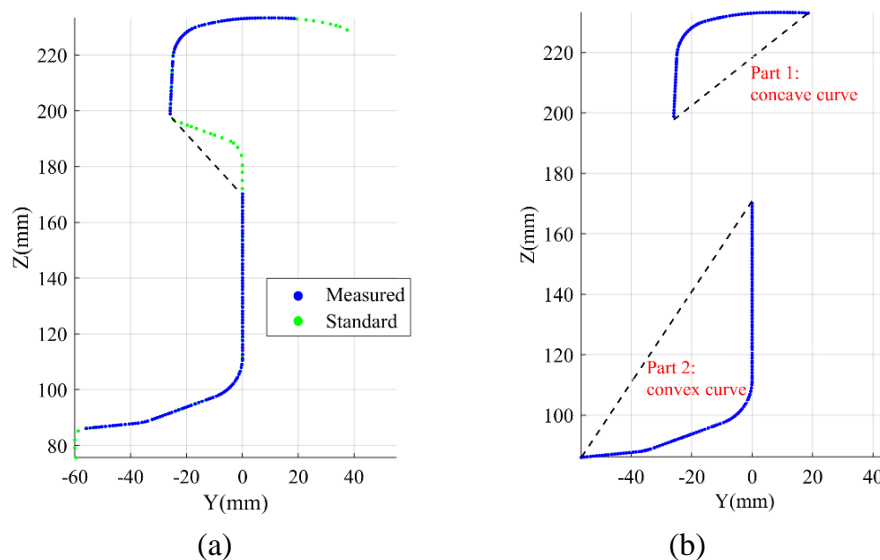
Fig. 2.23 Results of existing rail profile measurement: (a) interpolating between measurements; (b) dividing a discontinuous profile into two continuous parts

As an inspection technique, dedicated inspection methods or strategies are as important as inspection systems. Existing laser-based rail inspection methods are generally based on comparison with 2D profiles. Although there have been attempts to measure the 3D point cloud of the rail surface, separate data points cannot fully recover the geometrical characteristics of the rail surface where the points originate. As such, strategies that can be derived from 3D data points for inspection purposes are still limited. Overall, from an inspection method point of view, existing laser sensor data-based inspection methods can be summarised as follows:

1) **Diversity:** The diversity of inspection is low, as it is normally based on simple comparisons or thresholding methods. These methods are difficult to satisfy complex inspection cases in the real world. For example, it is difficult to choose the range of tolerance for a comparison-based method due to varying measuring conditions. Meanwhile, the standard profile or CAD may not always be available in practice. Similarly, artificial thresholds are impossible to cover complex situations in the real world, such as defects in a wide variety of sizes.

2) **Dimension:** Existing laser-based rail inspection methods are mostly implemented on a 2D basis. 2D rail profiles provide limited information for inspection. For example, the geometrical characteristics of particular rail surface defects such as surface cracks may be longitudinal in nature which are difficult to detect with 2D profiles. Therefore, it is worth considering increasing the dimension of the laser-based inspection method. The author's previous research explored the feasibility of reconstructing the railhead surface in 3D with the alignment of 2D profiles and a dedicated 3D reconstruction pipeline[10]. This will be further discussed and investigated in this research.

Laser triangulation sensors have their strengths in nature for surface measurement and inspection. However, their potential for rail inspection has yet to be fully realised. This subsection has analysed the limitations of laser-based rail inspection approaches from two points of view (the system and the inspection method), which provide possible research routes to improve its overall performance as a rail inspection approach.

2.4 Multi-sensor technologies

The previous subsections (2.2 & 2.3) discuss the strengths and limitations of existing vision and laser-based rail inspection approaches. After the analyses, an expectation would be how to make use of their strengths and eliminate the existing limitations to improve their overall performance. One of the potential solutions to achieve this would be multi-sensor technology. Generally, multi-sensor technology is defined as a process of applying a range of disparate sensors in a single set-up, to allow the dimensions of the target to be measured [75]. This process normally includes two components: multi-sensor integration and data fusion. The former is the combination of multiple sensors at the hardware level either stepwise or simultaneously, while the latter occurs at the data level, combining different sources of sensory data to achieve one representational

format. This provides a route to combine the strengths of different sensors and gathers information more efficiently and accurately than when the sensors are used individually.

In recent years, the development of sensory technologies and constantly improving processors have paved the way for applying multi-sensor technology in more areas. Examples of vision- and laser-related multi-sensor integration can be found in the fields of autonomous navigation and reverse engineering [76] [77], where vision and laser scanners are used for perception and visualisation of environments. The common challenge is the limited perception range of scanners so that they cannot cover the ROI from one location. One of the popular solutions, known as simultaneous localisation and mapping (SLAM), is to integrate vision and laser scanners with navigation sensors such as a global positioning system (GPS) unit and an inertial measurement unit (IMU), to track the trajectory of the system so that the scenes captured from different positions can be registered, yielding a complete scene [78]. In the field of rail inspection using NDT methods, multi-sensing technologies also commonly exist. For example, the robotic inspection system developed by Rowshandel combines ACFM with a 1D laser sensor for accurate surface defect detection and positioning [3]. The PMS developed by MERMEC integrates laser sensors with a robust IMU to track the dynamic behaviour of the vehicle for accurate rail profile measurement [6].

Lessons learnt from a variety of fields of applications indicate that multi-sensor technologies can normally offer increased system capacity and reduced measurement uncertainty. Meanwhile, new multi-sensor systems tend to be more compact and highly integrated to carry out more complex tasks with reduced human involvement. However, as two prominent sensing techniques, laser and vision sensors are mostly applied individually in the field of rail inspection, which limits the capability of the railway inspection system they represent, as indicated in sections 2.2 and 2.3. Therefore, the development of a multi-sensing system based on laser, vision, and other NDT sensors provides a great opportunity to combine their respective strengths to improve the overall performance of laser- and vision-based rail inspection.

2.5 Conclusions

Rail inspection plays a significant role in ensuring that the rail in use meets the relevant quality standards, and thus ensures the safe transit of the train and the normal operation of the railway network. Among all rail inspection tasks, the inspection of rail surface

defects caused by the long-term rolling contact between the rail and wheels is critical, as surface defects can directly lead to non-optimal contact with train wheels and have also been thought one of the principal predisposing causes of some of the more severe rail failures. Therefore, inspection of surface defects should aim to be accurate and efficient to support appropriate maintenance works and prevent further deterioration or accidents. Based on the relevant standard, this chapter firstly introduced the categorisation of rail surface defects and the characteristics of the defects in each category. Existing inspection solutions based on different NDT techniques were then reviewed. In comparison with conventional manual inspection methods, the application of these NDT techniques has considerably improved the efficiency of rail inspection. Among the existing NDT techniques, vision and laser techniques outperform others, benefiting from their non-contact and optical imaging properties. Accordingly, existing vision- and laser-based rail inspection approaches are reviewed and discussed in detail, through which their strengths and limitations as two separate techniques are further analysed. Since vision and laser techniques are all based on optical reflections, some of their strengths and limitations are complementary to each other. For example, the performance of vision-based systems is normally dependent on surface cleanliness and illumination equality, while laser-based systems are more robust to complex surface conditions and ambient lighting variations. Similarly, vision-based inspection strategies are diverse, with numerous inspection algorithms available, while laser-based inspection methods show low diversity. These observations suggested the route to combine the strengths of different sensing techniques and eliminate their shortcomings to improve their overall performance for rail inspection. This process is also known as multi-sensor technology. Accordingly, multi-sensor technology was introduced with related applications in other fields reviewed. Motivated by this, the development of a multi-sensing system based on laser, vision, and other NDT sensors was proposed, which aims to improve the performance of laser- and vision-based rail inspection in terms of both accuracy and efficiency.

3 PRINCIPLES OF VISION-BASED METHODS AND THEIR APPLICATIONS

3.1 Overview

The strengths and weaknesses of vision-based rail inspection through their applications in existing systems are reviewed in the previous chapter (section 2.2). To further investigate the performance of vision-based rail inspection, this section introduces a vision-based switch rail condition monitoring and inspection system as a case study. This case study is also a part of an industrial project that the author participated in (China High-Speed Railway: Hefei–Anqing 2019(001)), which aims to monitor the status of the switch rail and inspect switch rail faults, to ensure the safe operation of S&C and further support management of their maintenance.

3.2 Case study – A line-side switch rail condition monitoring and inspection system

As introduced in section 2.1.1, switch rails are one of the key components of S&C. They are controlled by switch machines to guide trains from one route to another. The discontinuity in the geometry of switch rails can expose them to high-impact loads as train wheels pass through (Fig. 3.1). Even with the use of advanced alloys, local plastic deformation and abrasion can still be derived with such high-impact loads in the long term. Defective switch rails can lead to incorrect alignment with the stock rail, which can be characterised by: 1) the abnormal contact angle between the switch rail and the stock rail, should the contact angle between the two become too shallow for example, a flange climb derailment could occur, and 2) the gap between the stock rail and the closed switch rail, especially at the switch rail toes where train wheels travel from the stock rail onto the switch rail, known as a toe gap. Excessive toe gaps can further

(a)   (b)
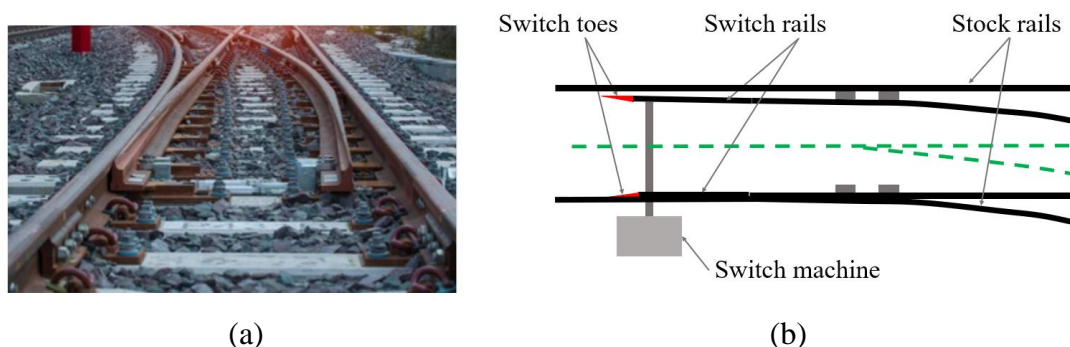
Fig. 3.1. Photo (a) and diagram (b) of the simplest structure of S&C

increase the impact loads as train wheels travel from the stock rail onto the switch rail with wheel flanges striking the switch rail. This can even make the switch 'crawl' along the stock rail in the running direction of the train, which endangers transit safety.

Traditionally, S&C inspection works are mainly reliant on a periodic visual inspection conducted by trained staff complying with corresponding standards, such as Network Rail's (NR) NR/L2/TRK/001 (particularly, sub-standards NR/L2/TRK/0053 and NR/L2/TRK/1054 for S&C) [79]. This process normally involves the use of one or several mechanical contact gauges. For example, the TGP8 gauge (Fig. 3.2 (a)) is commonly used by NR to identify the presence of a derailment hazard by checking the contact angle between the wheel and the switch rail [9]. Similarly, an S&C confirmation gauge (Fig. 3.2 (b)) is also used to check the distance between the switch and stock rails [80].

(a)
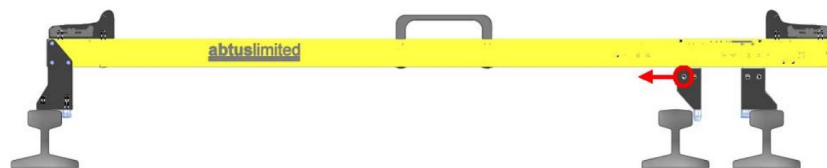
(b)

Fig. 3.2. Examples of contact gauges for S&C measurement: (a) NR TGP8 manual gauge; (b) Abtus ABT4670 confirmation gauge

However, these manual inspections are considered to be low-efficiency and susceptible to human errors. For example, the accuracy of manual inspections is normally difficult to define. Workers can adopt their own method of using the gauge, which can result in imprecise results, although there have been automatic inspection systems such as the

Unattended (track), Geometry Measurement Systems (UGMs) fitted on NR's NMT [4] and other systems with similar measuring units mounted on rail vehicles [7] [8] or trolleys [9]. These onboard systems need to be scheduled into the train timetable, and thus cannot support continuous or frequent condition monitoring. Therefore, there is still a risk of missing developing faults or deterioration of S&C. Moreover, the inspection frequency varies between railway administrations, traffic types, and line speeds. It is often very difficult to determine an appropriate inspection frequency. Accordingly, a vision-based switch rail condition monitoring prototype was proposed under such a context. The system aimed to monitor the status of the switch rail, including its movement, position, toe gap, and the edge of the toes in a real-time manner. The following subsections 3.2.1 and 3.2.2 introduce the system and related inspection algorithms, respectively. Section 3.3 further evaluates the strengths and weaknesses of vision-based rail inspection methods based on the field test results of the proposed system.

3.2.1   Data acquisition system

As discussed in section 2.2, vision-based inspection or monitoring systems normally consist of a data acquisition module and computer vision algorithms according to specific vision tasks. The data acquisition system in this case study uses a pair of high-resolution CCD cameras to capture the images of the switch rail. Additionally, the system contains a pair of external infrared (IR) lights to ensure the system's imaging quality when the ambient light dims at certain times of day, and an embedded PC to communicate with a remote server and transmit the data for further processing.

To monitor the status of the switch rail and have minimal impact on the existing systems, the system was expected to be installed at the trackside or on the overhead line. The proposed system in this case study was fitted at the trackside as shown in Fig. 3.3 (A and B). This installation allows for the visibility of both toe gaps and the edge of switch rail toes. The dashed lines indicate their targeted areas (switch rail). The camera module contains a zoom lens and a focusing system which ensures high-quality images while also offering the ability to focus tightly on the target areas from different distances. At the testing site where the tests were conducted (Hefei, China), the height of the system was 0.55 m and its relative distance to the midpoint between the stock rails was 2.40 m, to satisfy the safety requirements for line-side equipment [79]. The system can output

30 frames per second at a resolution of 1920×1080 during operation, which ensures continuous monitoring of the target objects.
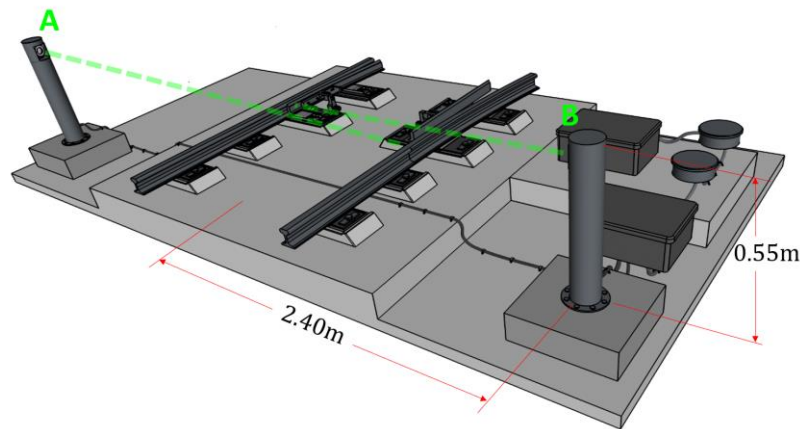
Fig. 3.3. CAD model of the proposed system

3.2.2 Methodology

To measure the toe gap and monitor the status of the switch rail, the vision-based system in this case study must be able to detect and locate the switch rail in consecutive video frames. The status of the switch rail mainly refers to the edge of the switch rail toe where it approaches the stock rail. Any abrasion or wear at this point can cause non-optimal contact with the wheel flange and may also distort the toe gap. Therefore, the edge of the switch rail toe needs to be identified when locating the switch rail. These activities yield two main vision tasks: 1) detection of the switch rail, and 2) detection and evaluation of the switch rail edge, which should be considered in detail to understand the performance and characteristics of vision-based detection methods in actual applications. It is worth noting that some processes such as stock rail localisation are not discussed as they are not related to this research. Details of the complete process can be found in the published paper [81].

(1) Detection of the switch rail

For a vision-based system, object detection is the process of locating an object in an image based on its discriminative features. Similar to the feature extraction methods discussed in section 2.2.2 (where the object is the defect), the features for the detection of the switch rail could be its colours, edges, and corners, etc. These features can be further described as quantitative attributes by using feature descriptors such as Speeded Up Robust Features (SURF) and Maximally Stable Extremal Regions (MSER). In the

initial tests, SURF was used as a descriptor to locate the switch rail by detecting the locking nut on the driving rod. SURF focuses on the points of interest of the object in the image such as edges and corners with higher intensity values [82]. Fig. 3.4 illustrates the key stages of the SURF-based method. It works by finding the points of interest in both the reference image which contains only the object (Fig. 3.4 (a)) and the whole video frame (Fig. 3.4 (b)) by using the Hessian matrix and Haar wavelet responses. The object can then be located by finding the matched interest point between the reference image and the video frames.
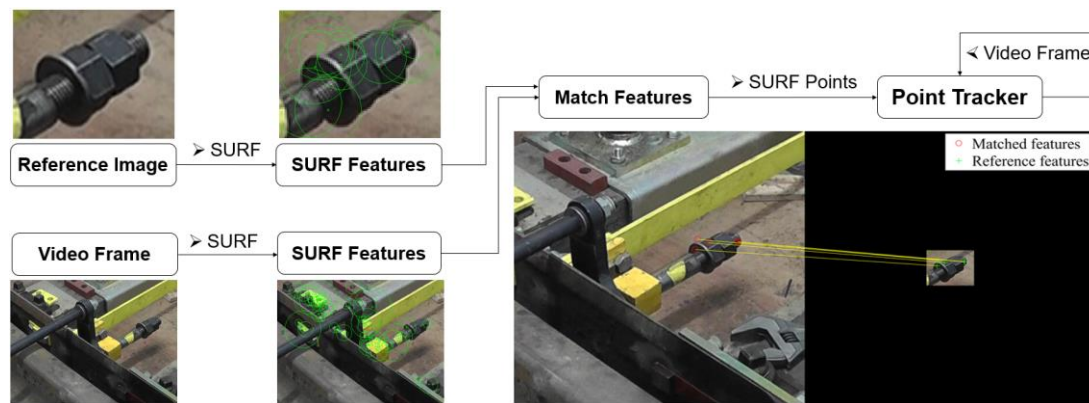
Fig. 3.4. SURF-based switch rail detection

During the field test stage, it was found that similar descriptors based on the surface texture features of the switch rail or its components can perform well in the lab where the working conditions such as backgrounds and lighting sources are stable. However, switch rails in service normally operate in varying conditions such as different weather conditions, backgrounds, and different S&C types. These factors can affect both the stability and accuracy of detection based on surface texture features and make this challenging for final deployment. Under this circumstance, one alternative solution that has been extensively used in other fields such as surgical instrument tracking [83] and augmented reality to assist the detection of the object [84] is the application of marker points to assist detection of the object. Fig. 3.5 shows two video frames captured by the system at the testing centre in the daytime (a) and at night (b). The marker point (in the dashed box) attached at the front of the switch rail is a round IR-reflective patch (30 mm diameter) with a black contour to facilitate detection with increased contrast against the background. The use of the IR-reflective material makes the marker point distinct against the background, especially at night when the external light dims, and detection of the switch rail can be converted to detection of the marker point. In the final stages

of this case study, such a reference marker point-based method had to be adopted to ensure the effectiveness of the system.

<div align="center">(a)   (b)</div>

Fig. 3.5. Testing switch rail with a marker point attached (a) daytime (b) night

To detect the marker point, and thus locate the switch rail, two strategies were considered. One is to detect the marker point by its circular shape using shape-based detectors such as Circular Hough Transform (CHT). However, as a moving object in the image, the shape of the marker point can vary in the form of scale changes. Moreover, the trajectory of the switch rail movement is not in line with the perspective of the camera. These can affect the accuracy of shape-based detectors and cause problems such as centre point deviation due to inaccurate marker point detection. Another strategy is to detect the marker point based on its intensity. The intensity is related to the brightness of the pixels in the image. It was observed that when the image is analysed in HSV (Hue, Saturation, Value) channels, pixels corresponding to the marker point have a relatively fixed intensity distribution in the V-channel (V-channel independently represents the brightness of each pixel). Furthermore, this property is less affected by external factors such as background changes and light fluctuations. Therefore, the intensity-based strategy was used.

Fig. 3.6 shows the framework of the method using two samples of video frames captured from the testing site. The initial video frames are firstly converted to HSV format (Fig. 3.6 (a)), where the intensity analysis of the V-channel is carried out (Fig. 3.6 (b)). The red vertical lines indicate the range of intensity values of pixels corresponding to the marker point. As shown in the partially enlarged views in Fig. 3.6 (b), clear peak and valley areas can be seen to distinguish these pixels from others. The experimental results demonstrated that the intensity values of pixels corresponding to the marker point have a stable range from 0.85 to 1.0, where 1.0 represents the
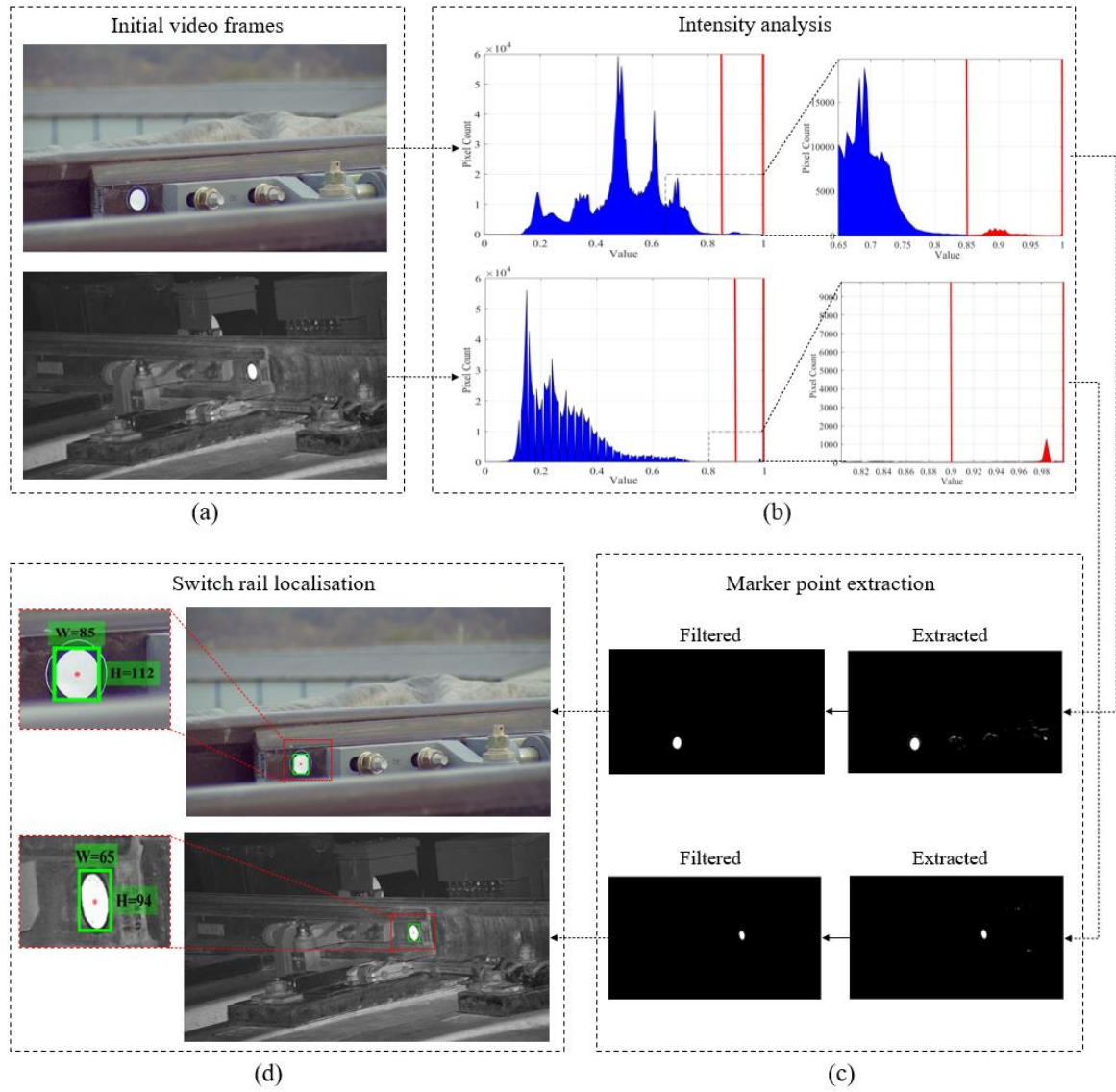
Fig. 3.6. Framework of marker point-based switch rail detection: (a) initial video frames captured in the daytime and at night; (b) V-channel intensity analysis and partial enlarged views; (c) detection results in binary images; (d) switch rail localisation

maximum intensity value of the image. Accordingly, an adaptive threshold can be defined by (3-1):

$$V_{thresholding} = 0.85 \times Max(V) \qquad (3-1)$$

where $Max(V)$ represents the maximum intensity value of the image in the V-channel. This process helps to detect the marker point in the image by replacing all the pixels above the global threshold ($V_{thresholding}$) with '1' and others with '0', yielding binary

images as shown in Fig. 3.6 (c). Randomly distributed noisy points can be filtered by using morphological operations such as erosion and dilation [85].

The marker point can then be localised in the image with a bounding box. The pixel size of the bounding box and the actual size of the marker point (30 mm in this application) can reflect the theoretical resolution of the detection and help to convert pixel distance to distance in millimetres. For example, one of the marker points in Fig. 3.6 (d) is localised with a bounding box of 65 pixels in width and 94 pixels in height. This means that the resolution of each pixel in this frame is about 0.46 mm (30/65 mm) in the longitudinal direction and 0.32 mm (30/94 mm) in the vertical direction.

(2) Switch rail edge detection and evaluation

Switch rail edge detection aims to detect the edge of the switch rail toe where the train wheels travel from the stock rail to the switch rail. There are many candidates such as Sobel, Canny, Prewitt, and Roberts detectors, all of which can be used as edge detectors [86]. Edge detectors are normally based on the gradient of the brightness of neighbouring pixels. In this application, the Canny edge detector performs better than the others; it calculates the brightness gradient of neighbouring pixels using the derivative of a Gaussian filter, which makes it relatively more immune to noise. Fig. 3.7 illustrates the framework of the edge detection method used in this case study. Firstly, the existence of the marker point makes it possible to locate the switch rail toe by cropping the ROI. This helps to initially filter out unrelated backgrounds to improve the detection accuracy and also reduces the computational complexity. A grey equalisation defined by (3-2) is then applied to the ROI to eliminate the noise and increase the contrast.

$$C(y,z) = \frac{R(y,z) - g(y)}{R(y,z) + g(y)} \tag{3-2}$$

where $R(y,z)$ represents the intensity value of the ROI, $g(y)$ is the average greyscale value on the y-axis of the ROI, and $C(y,z)$ is the equalised greyscale image.

The greyscale image of the ROI is then fed into the edge detector. In the Canny edge detector, the selection of a pair of intensity gradient thresholds determines the accuracy of the detection. As shown in Fig. 3.8, the thresholds are defined by two values, low and high. The rule of Canny is that any pixels with an intensity gradient larger than the
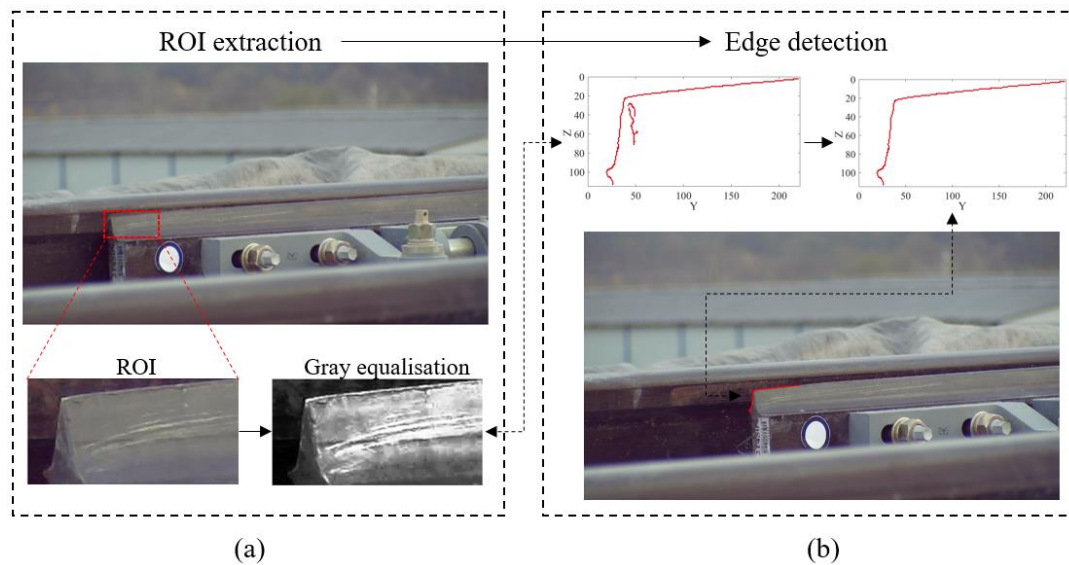
Fig. 3.7. Framework of edge detection and evaluation: (a) ROI extraction and grey equalisation; (b) edge detection of the switch rail toe; (c) affine registration for edge evaluation
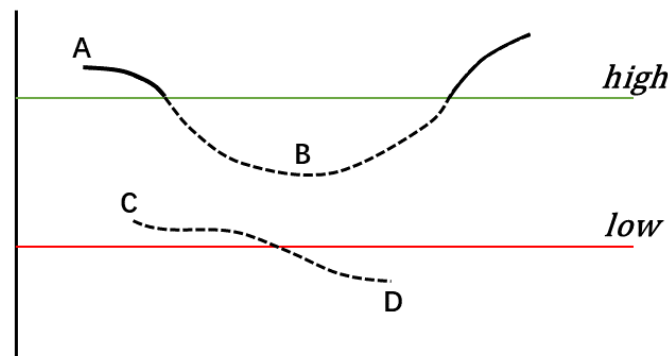
Fig. 3.8. Adaptive thresholding of Canny edge detector

high value are sure to be edges, and those below the low value are ignored as non-edges. Pixels with an intensity gradient lying between the thresholds are classified based on their connectivity, i.e., pixels indicated as B are also considered as a valid edge as they are connected to pixels above the high value.

For a single image, a pair of fixed thresholds is enough. However, the edge of the switch rail toe in this case study needs to be identified in consecutive frames. This brings a problem that the intensity values of the switch rail can change, especially along with the variation of the ambient light. In this case study, the Canny edge detector was optimised with a pair of adaptive low and high thresholds defined in (3-3):

$$low = (1 - \sigma) \times M$$

(3-3)

$$high = (1 + \sigma) \times M$$

where σ is an index of the sensitivity of the detector and can be a value from 0 to 1, and a larger σ indicates that more edges will be detected; M represents the median of the image intensity which helps the range of the thresholds adapt to the change of illumination conditions.

Fig. 3.7 (b) shows the detection results on a sample frame. Noisy points with random distribution and connectivity can be removed using morphological operations. The detected pixels can then be highlighted in the video frame, indicating the status of the switch rail in real time.

3.3 Field tests and evaluation

The vision-based system in this case study was tested in a testing centre for high-speed trains in China. Fig. 3.9 (a) shows a photo of the final deployment of the system (A and B). Fig. 3.9 (b) shows a monitoring result. The vision system used in this case study can monitor the status of the switch rail including the position, the gap between the switch and stock rails, and the edge of the switch rail toe. Continuous monitoring ensures the safe operation of the S&C and is also important to support the management of its maintenance.

(a)   (b)
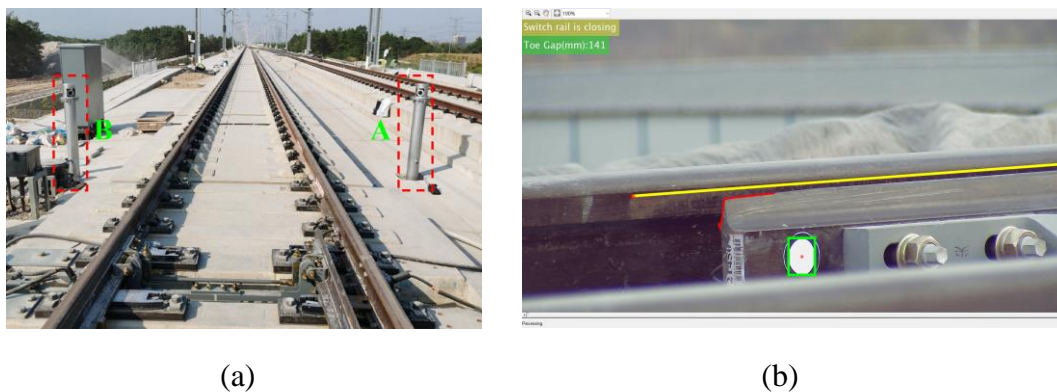
Fig. 3.9. Field test results: (a) completed prototype of the system; (b) monitoring results

In section 2.2.3, it has been discussed that the performance of a vision-based inspection system can be affected by the complicated working conditions in the real world. To further investigate these external factors and their impact on the system, the vision-

based system in this case study was also tested under various working conditions. Fig. 3.10 shows some representative results including tests conducted in different ambient lighting conditions and simulations of excessive toe gaps by putting obstacles between the switch and stock rails. For clarity, the working conditions and testing results corresponding to each sub-figure are summarised in Table 3.1. It can be observed from Fig. 3.10 (b) and (d) that the measurement of toe gaps based on the localisation of the marker point is accurate in most cases due to its higher intensity and the use of external IR lights at night. However, the edge detection is easily distorted (Fig. 3.10 (b) and (c)) or even invalid (Fig. 3.10 (d)) when the ambient light dims or at night. This is mainly due to the partial shadowing on the switch rail leading to poor contrast against the background.

(a)

(b)

(c)

(d)
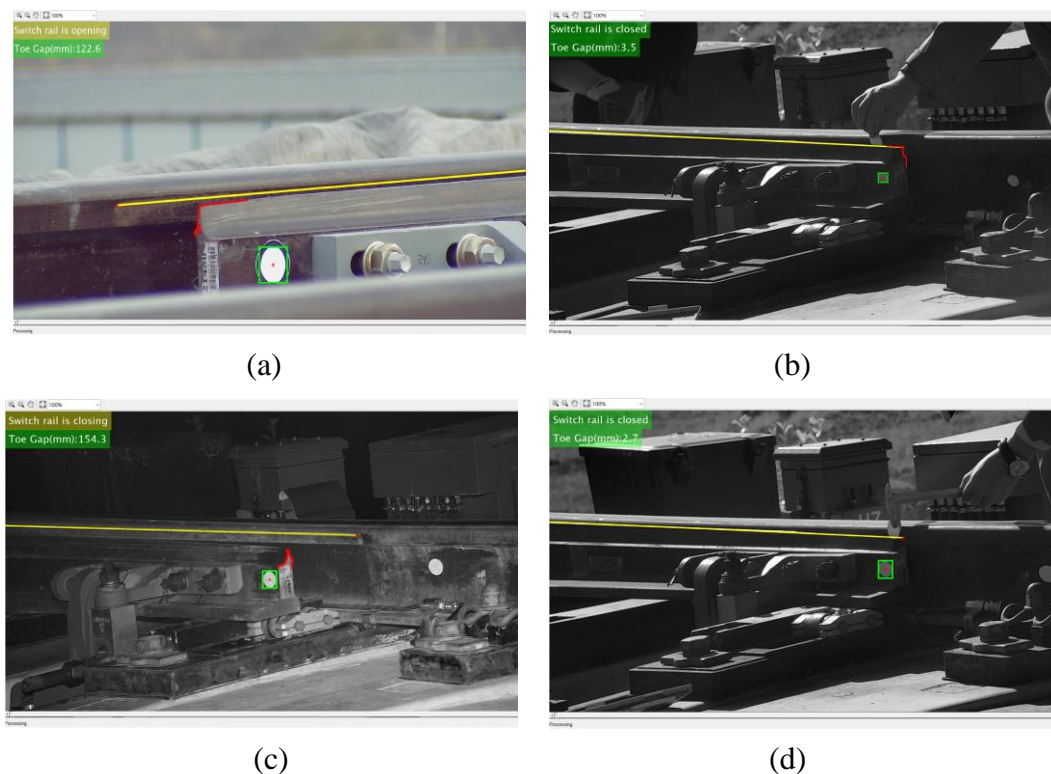
Fig. 3.10 Tests in different external conditions

Table 3.1 Working conditions and testing results indicated in Fig. 3.10

Fig. 3.10	Lighting condition	Toe gap (mm)		Edge detection
		Measured	Actual	
(a)	Morning /Normal	-	-	Valid
(b)	Dim	3.5	3.0	Distorted
(c)	Night	-	-	Distorted
(d)	Dim	2.7	2.0	Invalid

Further accuracy and stability tests demonstrated that the localisation of the switch rail performs well under consistent conditions. Fig. 3.11 shows a group of localisation results for the switch rail in its locked and unlocked positions, in which the green lines are the correct results. The blue line indicates the system localisation results giving an average error under 1 mm with a standard deviation of about 0.3 mm (indicated by red dotted lines). These results indicate that vision-based object detection is normally dependent on distinct surface texture features and sufficient external lighting conditions. However, these are normally difficult to achieve in practical applications with complicated working conditions. In these situations, some supporting measures such as the use of marker points, as demonstrated in this case study, may be necessary to ensure the system's effectiveness.
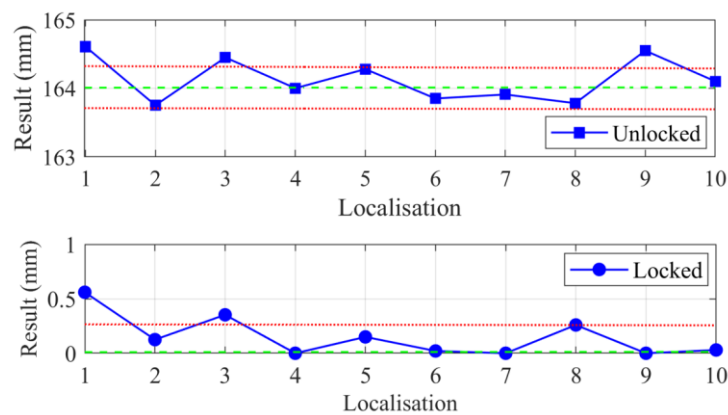
Fig. 3.11. System accuracy and stability analysis

Apart from the lighting conditions discussed above, many other factors could also affect the performance of a vision-based rail track inspection system. It is difficult to consider all of them in one case study due to experimental constraints, but it is worth discussing these for future reference.

1) **Surface cleanliness** is one of the important requirements for most optical inspection systems. For vision-based rail inspection, surface dirt on the rail is normally a major concern as it may cause inaccurate detection results, and such augmentations on the camera lens may even affect its function.

2) **Adverse weather conditions** such as fog, rain, or snow happen in the outdoor environment. In these cases, the object may be partially obscured, which can affect the system's imaging quality and lead to noisy and distorted detection results.

3) **Material degradation** in rail inspection mainly refers to degradation of the rail due to the long-term wheel–rail contact interface. For example, in this case study, the running surface of the switch rail may develop complex profiles, which may affect the accuracy of edge detection. Similarly, the running surface of the rail can be reflective and develop varying surface textures. These can also affect the effectiveness of vision-based detection algorithms.

Given these external factors, additional measures may need to be taken to ensure the effectiveness of vision-based systems. For example, surface dirt on the rail is normally related to the application of grease. Adhesive grease accelerates the accumulation of surface contaminations. Therefore, at the point where a vision-based system is in operation, the target object of the system should not be covered by grease. Meanwhile, key components such as the switch rail and the camera lens in this case study are expected to be cleaned periodically to further remain surface cleanliness. In this case study, the use of reflective marker points and external IR lights increases the visibility of the object, and have been proven to be effective in dealing with some adverse weather conditions such as fog. However, other measures such as adding a screen wiper on the camera lens, combining other sensory techniques such as the laser as will be discussed later in this research, or using specific artefact detection and removal algorithms may need to be considered if more adverse effects (such as adherent raindrops or snow) occur. Material degradation can affect the effectiveness of computer vision algorithms. Therefore, for an application involving material degradation, vision-based algorithms may need to be calibrated periodically according to actual conditions.

3.4 Summary

This chapter aims to identify the strengths and weaknesses of vision-based railway track inspection techniques by choosing a project of vision-based switch rail condition monitoring and inspection as a case study. Through the case study, the main strengths of vision-based systems can be summarised in two points: 1) they are easy to deploy and can be non-destructive in most cases, and 2) numerous vision-based detection algorithms that have been developed and optimised are available, and thus can fit different vision tasks. However, the weaknesses of vision-based inspection techniques are also significant. Firstly, the fundamental point of vision-based inspection algorithms is normally the distinguishable texture features of the object against backgrounds.

However, this is susceptible to problems in practical applications. For example, railway tracks can be dark and oxidised in practice which can affect their surface texture features. In this case study, a reference marker point-based method had to be adopted to ensure the stable detection and localisation of the switch rail. Secondly, as an optical imaging technique based on passive light reflections, vision-based systems normally require relatively good and consistent lighting conditions to ensure imaging quality. However, the lighting conditions of the outdoor environment can vary throughout the day. Even with the use of external lighting sources, the imaging quality can still be affected by adverse weather conditions which could, in turn, affect the accuracy of vision algorithms such as the edge detection algorithm in this case study. Therefore, the accuracy and stability of vision-based inspection systems in practical applications are often questioned. To ensure the system's effectiveness, some precautions or additional measures should always be considered to support its application in specific tasks.

4 PRINCIPLES OF LASER-BASED METHODS AND THEIR APPLICATIONS

4.1 Overview

High-precision laser triangulation sensors are among the most popular optical measurement instruments that can be applied for surface defect inspection. Compared to vision-based methods, laser sensors access surface geometries by actively projecting laser lights and observing the diffusely reflected laser light. Modern laser sensors can generate high-resolution, high-energy, and structured laser lights with the use of highly collimated light sources and special lenses. Therefore, laser sensors are naturally more accurate and show more tolerance to external factors such as complex surface conditions compared to vision sensors that are reliant on passive natural light reflections. In section 2.3, some of the laser-based systems are reviewed, with some of their strengths and weakness analysed. Existing systems normally use 2D laser triangulation sensors as onboard measuring units and use the linear nature of the rail to measure 2D rail profiles with the movement of the train or the mobile platform. This section goes into more depth by using a 2D laser triangulation sensor-based prototype system developed by the author as a case study to further identify the characteristics of laser-based rail inspection. Through numerous laboratory and field tests, the characteristics and potentials of laser-based rail inspection were explored. The author has also published this part of the work in [10], some of which is worth discussing here.

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

Most laser-based rail inspection is carried out in the form of 2D rail profile measurement. However, the information that 2D profiles can provide is limited, which limits both the dimension of measurement and the diversity of inspection methods. This has been initially analysed in section 2.3.3. In recent years, with the development of sensory technology, laser-based 3D equipment is becoming available. Some attempts to use laser-based 3D imaging systems for rail inspection have also been introduced in section 2.3.1. It has been summarised that 3D laser sensors have the advantage regarding the dimension of measurement. However, they still have limitations in terms of the cost of the scanner itself and the measurement accuracy and efficiency. Therefore, the case study introduced in this chapter still used a mainstream 2D laser triangulation

sensor, while the feasibility of high-precision 3D rail measurement based on such a low-cost 2D laser triangulation sensor was also explored to help comprehensively identify the characteristics of laser-based rail inspection.

4.2.1 Data acquisition system

As has been mentioned, existing laser-based rail inspection systems normally use 2D laser triangulation sensors as the main measuring units. To mimic the application scenarios where lasers are commonly used, a simple four-footed mechanical frame was built in the lab (Fig. 4.1 (a)). The frame contains two slide rails in both longitudinal and lateral directions as the main operating platform of the laser. The laser module used in this case study was a scanCONTROL 2900-100 blue laser sensor from Micro-Epsilon owing to its properties of compact physical size, high resolution (1280 points/profile), and high frequency (up to 2000 Hz) [87]. To hold the laser sensor, a mechanical arm with six degrees of freedom (6DoF) was used. Apart from the laser sensor, two draw-wire displacement sensors were mounted on the slide rails to track the position of the
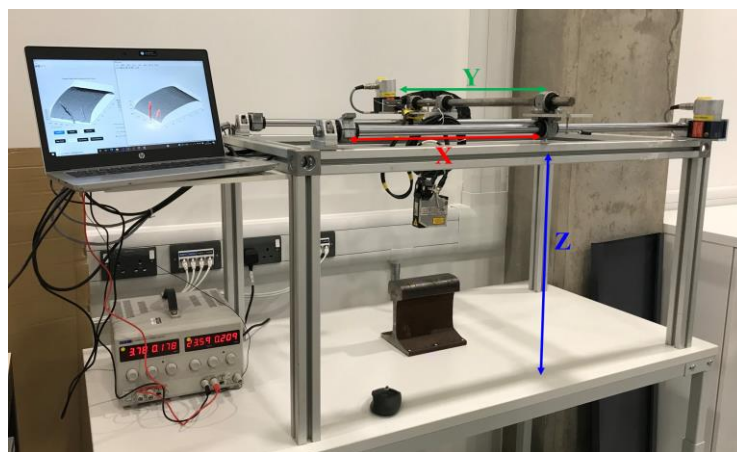
(a)
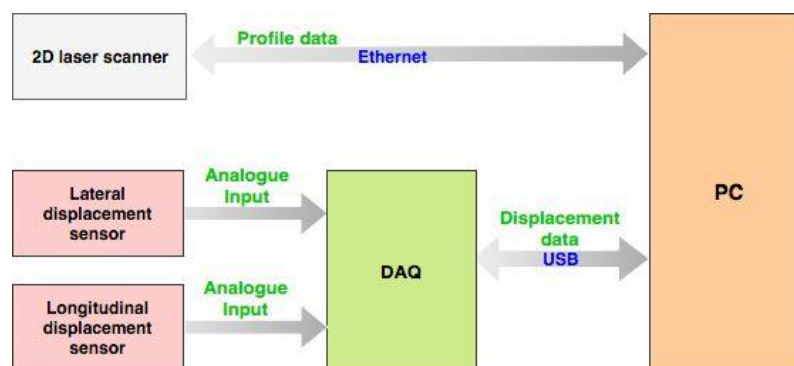
(b)

Fig. 4.1. Laser-based rail inspection: (a) completed prototype in the lab; (b) framework to the system

laser sensor. This deployment allows the laser to be moved above the rail and pointed at any area of the rail to measure 2D rail profiles. Meanwhile, 2D profiles can be aligned in a global coordinate system according to the corresponding position of the laser sensor at the time of scanning to increase the dimension of the measurement. Fig. 4.1 (b) shows the framework of the system and the data flow of the main components. The digital laser sensor data is transmitted to the PC using Gigabit Ethernet as the standard connection. The analogue data from the draw-wire sensors are transmitted to a data acquisition module (DAQ: NI USB-6210) containing 16-bit ADC converters. The digitised displacement data are then fed into the PC synchronously.

The final integrated system can output 2D profiles with a resolution of up to 1280 points per profile. The displacement sensors measure the linear movement in terms of analogue voltage changes. Therefore, their resolution is determined by the smallest incremental voltage that can be recognised. In this case, the reference voltage of displacement sensors is from 0 to 10 V representing the physical displacement from 0 to 1500 mm. Therefore, the resolution of the system in the lateral and longitudinal directions can be calculated using (4-1) and (4-2).

$$\frac{10 \ V}{2^{16}} = \ 152.5 \ \pi V \tag{4-1}$$

$$\frac{1500 \ mm}{2^{16}} = \ 0.023 \ mm \tag{4-2}$$

where 2^{16} is the number of digital values that a 16-bit ADC converter can provide. Accordingly, each 2D profile contains 1280 points, and the smallest interval between each profile is 0.023mm. The resolution of the integrated system is therefore considered to be accurate enough to characterise rail surface conditions.

4.2.2   Methodology

With the platform introduced above, this case study mainly considered two application scenarios of the 2D laser sensor. One is *2D rail profile measurement* as most of the existing systems do, and the other is *3D measurement* that allows for more complete geometric information by aligning 2D rail profiles along the longitudinal direction of the rail. Fig. 4.2 illustrates the specific region of the rail covered by the laser sensor in this case study. As has been discussed in section 2.1.1, the surface defects that this research focuses on are mostly distributed on the rail head (including the gauge corner

region, running surface, and field corner region as illustrated in Fig. 4.2 (a)), where train wheels make contact; the laser sensor was deployed vertically on the top of the rail to ensure maximum coverage of the ROI (Fig. 4.2 (b)).

(a)   (b)
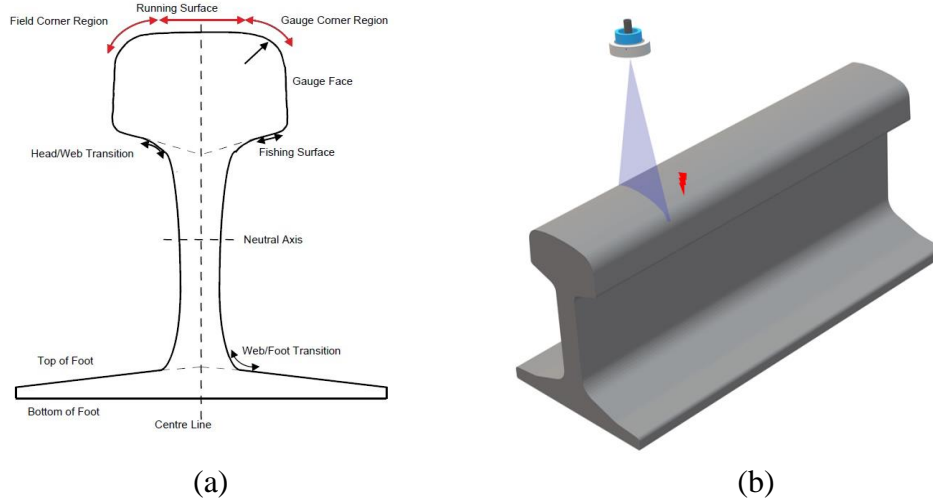
Fig. 4.2. Region of interest in this case study: (a) cross-section profile of plain track; (b) deployment of the laser sensor

(1) 2D measurement

The system in this case study can output 2D rail profiles directly with a frequency of up to 2000 Hz. Fig. 4.3 shows some examples of 2D measurement results. Among the sub-figures, (a) and (c) are the measurement results, and (b) and (d) are the processed cross-section profiles of the sample with key features of defects highlighted. It can be observed that, in 2D profiles, surface defects normally appear with corresponding data points out of the normal plane. Most of these changes can be captured by the laser sensor. Apart from the comparison with corresponding standard profiles as mentioned in section 2.3.2, these defects are also detectable based on their geometrical characteristics. For example, Fig. 4.3 (a) shows the measurement result for a rail sample containing surface cracks. The region that corresponds to the defect can be detected by identifying data points with sharp changes in depth, as illustrated by its processed version in Fig. 4.3 (b). The fundamental of this method is that the depth of neighbouring points in the normal surface changes smoothly with a small gradient, while the depth of neighbouring points around the defect changes shapely with a relatively large gradient. In mathematics, this can be represented by (4-3):

$$|\nabla H(x_i, y_i)| = max(H(x_i, y_i)) - min(H(x_i, y_i)) \qquad (4\text{-}3)$$

where $max(H(x_i, y_i))$ and $min\,(H(x_i, y_i))$ represent the maximum and minimum depth of neighbouring points, respectively.

<div align="center">(a)</div> <div align="center">(b)</div>

<div align="center">(c)</div> <div align="center">(d)</div>
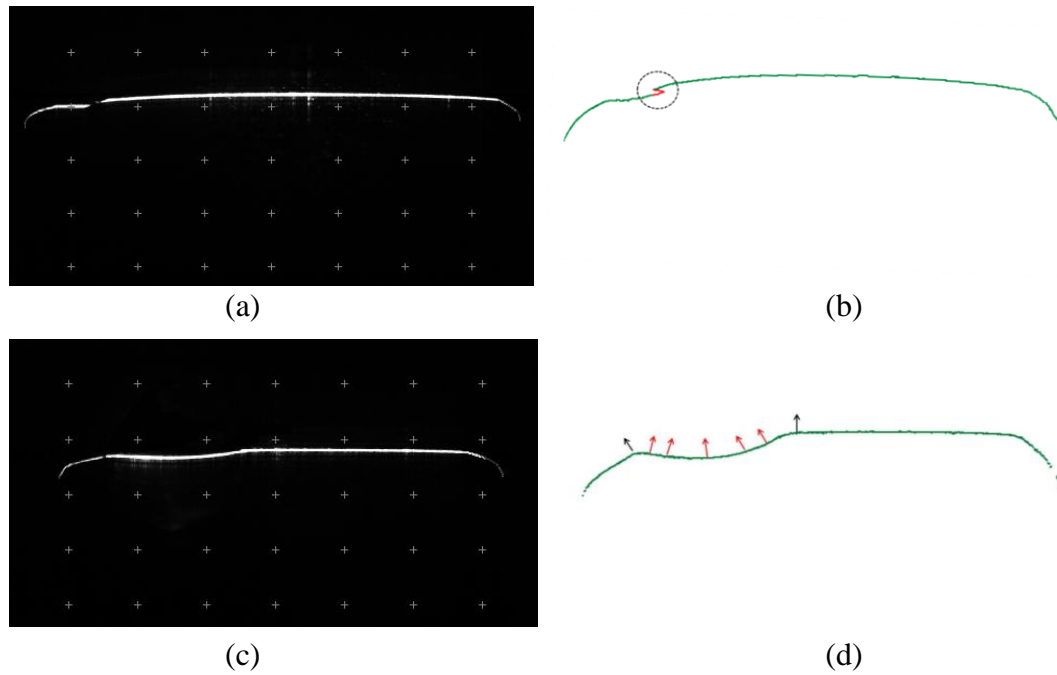
Fig. 4.3. Cross-section of rail tracks: (a) laser scanning result for a rail sample suffering from surface crack; (b) processed data from (a) with key features highlighted; (c) laser scanning result for a rail sample suffering from squats; (d) processed data from (c) with key features highlighted

Similarly, Fig. 4.3 (c) shows the measurement result for a rail sample containing surface squats. Such defects normally form a partially concave hull in the profile with sharp changes in surface normal vectors. Fig. 4.3 (d) indicates the distribution of normal vectors around the defect with arrows. The normal vectors at the edge of the defect are not in alignment with those on the normal rail surface. Mathematically, this can be represented by (4-4):

$$\left|\nabla N_j\right| = \max_{|v|=1}[\nabla_v N_j] = max(N_j) - min\,(N_j) \qquad (4\text{-}4)$$

where $max(N_j)$ and $min\,(N_j)$ represents the maximum and minimum normal vectors of points in a subset, respectively.

Through specific geometrical and thresholding analysis, defects can be detected with the use of 2D rail profiles. However, it is not difficult to see that the information that 2D profiles contain is limited. They only contain geometric information of the rail in

the lateral and vertical directions. More complete geometries of surface defects are longitudinal in nature. Therefore, more complicated inspection tasks such as defect classification and evaluation are difficult to achieve using 2D profiles, as they may show similar and indistinguishable geometries in 2D.

(2) 3D measurement

The platform used in this case study contains displacement sensors to track the movement of the laser sensor in the X–Y plane. This allows for the alignment of 2D profiles in the third dimension so that surface conditions can be represented more comprehensively with data points in 3D, also known as the point cloud in computer graphics [88], as shown in Fig. 4.4.
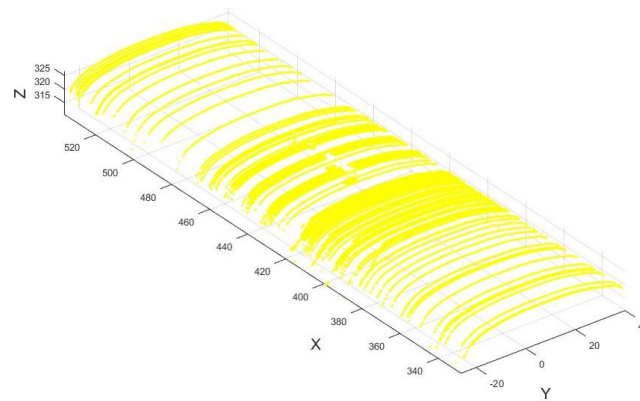
Fig. 4.4. 2D profiles in sequence

However, the manually controlled movement of the laser sensor in this case study determines that the distribution of 2D profiles in the X-axis is not uniform. Some regions have excessive point density while others have gaps with less point density. These increase the complexity for further analysis. In this case study, two pre-processing steps were used: (1) *down-sampling* and (2) *interpolation*. Down-sampling is an image processing technique for sampling raw image data into a lower resolution. Appropriate down-sampling methods can also be used here to remove redundant points in the point cloud data. The method used in this case is a Grid Average method [89], which firstly generates an axis-aligned bounding box for the entire point cloud. The bounding box is then divided into thousands of small grid boxes of a specified size. Raw data points within each grid box are merged into a single point by averaging their locations and normal vectors. Therefore, the density of down-sampled point clouds can be controlled by the size or the density of the grid box. In this case study, $1.0 \times 1.0 \times 1.0$ mm grid boxes were used. This ensures that the redundant points are

filtered while also preserving the geometry of the raw point cloud. Fig. 4.5 shows the down-sampled point cloud. The number of data points in this example is reduced from 25,600 to 8604, which considerably reduces the computational complexity for further analysis.
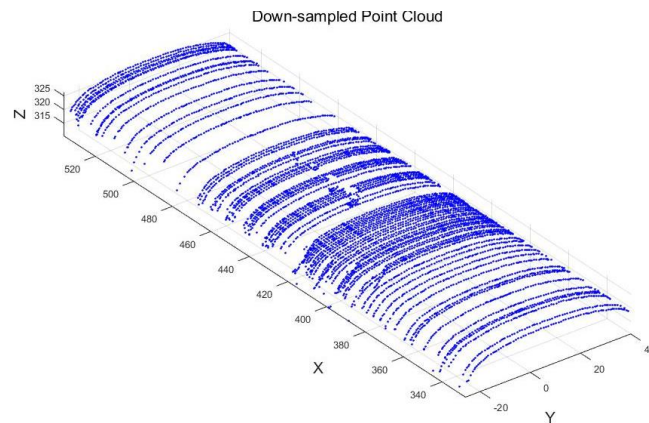
Fig. 4.5. Down-sampled point cloud

To accurately represent rail surface conditions, the point cloud data are expected to be structured and coherent in the 3D space without gaps or holes. However, the distribution of data points in the down-sampled point cloud is non-uniform, with gaps in the longitudinal direction of the rail. This is mainly due to the manually controlled movement of the laser sensor along the longitudinal direction of the rail in this case study. The laser sensor was moved at non-uniform speeds. For example, for the region which is defect-free, the surface needs a few points to define, so the laser sensor can be moved quickly. On the contrary, for the region in which defects are distributed, the laser sensor was moved at relatively slow speeds to allow for reliable characterisation of rough features with more points. To solve these issues, interpolation was applied on the down-sampled point cloud. Interpolation is a common technique in the area of 3D reconstruction, which generates an interpolating surface that encloses the raw data points in space. The interpolating surface can then be used to compute a set of interpolated points that preserve the geometry of the raw data points and are also structured in space. In this case study, a method called Delaunay triangulation was used among others to generate the interpolating surface. This method works by searching for triangles that fit raw data points, as illustrated in Fig. 4.6.

The interpolating surface (F) formed by these triangles can be simplified using (4-5):

$$Z = F(X, Y) \qquad\qquad (4\text{-}5)$$

where X and Y are the coordinates of the raw data points in the X–Y plane, and Z is the corresponding depth of the point on the Z-axis.
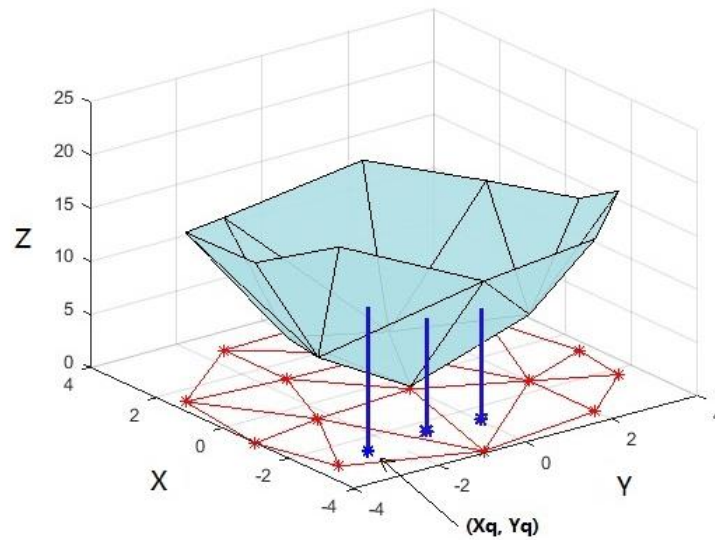
Fig. 4.6. Basics of Delaunay triangulation

Once the interpolating surface is generated (red connected points in the X–Y plane in Fig. 4.6), new data points (X_q, Y_q) that are distributed uniformly in the X–Y plane with a specified density can be interpolated (blue connected points in the X–Y plane in Fig. 4.6). Since the interpolation is also a process of approximation for filling gaps, a maximum triangle size of 20 mm was set to ensure the accuracy of the interpolating surface. This means that any gap and hole that is larger than 20 mm in diameter will be ignored. The depth Z_q of each point is derived from the depth of raw data points in the nearest triangle. Fig. 4.7 (a) shows the point cloud after interpolation. From the top view of the point cloud (Fig. 4.7 (b)), it can be seen that all the points are distributed uniformly. In this example, the density of the interpolated points is 220 (X) $\times$ 70 (Y). In other cases, the density can be adjusted according to the size and the structural complexity of the object.

(a)   (b)
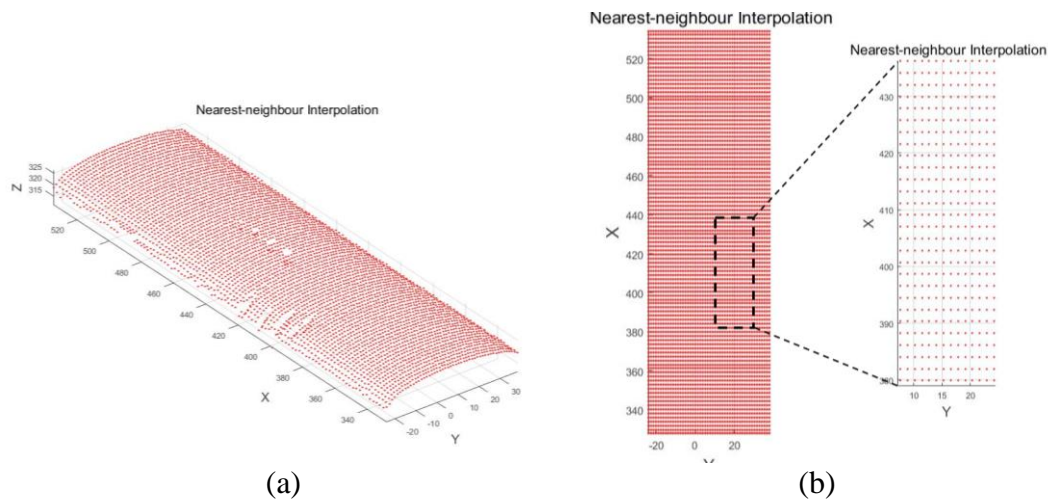
Fig. 4.7. Nearest-neighbour interpolation: (a) interpolated point cloud; (b) distribution of points in the X–Y plane

The interpolated point cloud represents the condition of the railhead surface with coherent data points, with which the geometrical properties can be recovered by rebuilding the surface. This can be done with surface polygon meshes, which connect scattered points with polygons [90]. Fig. 4.8 shows the polygon mesh result for this example. From its partially enlarged view, it can be seen that the geometry of the railhead surface is characterised by thousands of quadrilaterals. Compared to discrete 2D rail profiles, this 3D model characterises the railhead surface conditions more comprehensively, as it contains more information that can help to achieve automatic detection, classification, and evaluation of surface defects in future processing.
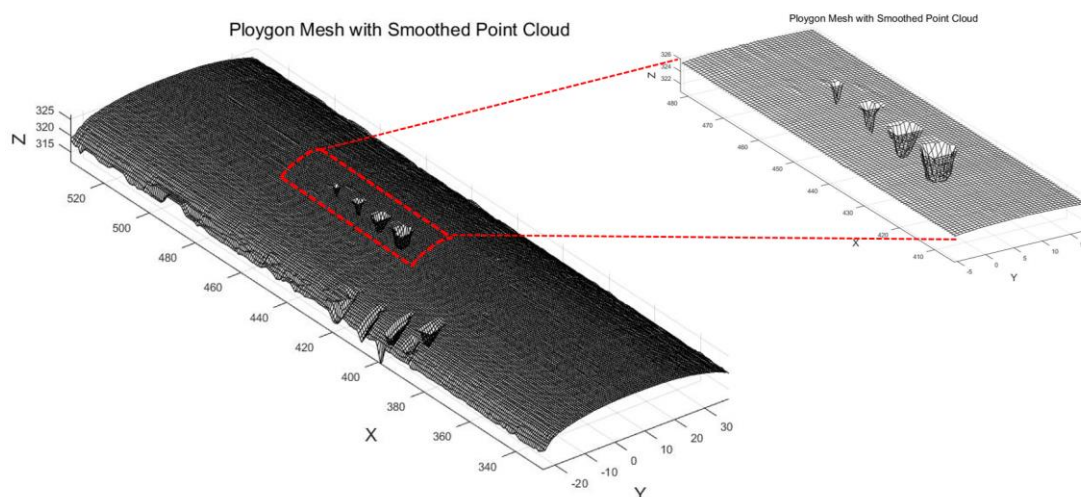
Fig. 4.8. Polygon meshed point cloud

4.3 Field tests and evaluation

Laser-based rail inspection shows promising results in the laboratory, especially for 3D measurement. To further evaluate the performance of the laser sensor and explore the feasibility of 3D measurement, more tests were carried out in both laboratory and real-world environments. Fig. 4.9 shows some samples that have been measured using the system proposed in this case study. These include rail samples that were machined and spoiled in the lab, and an in-service rail containing real defects and joints. Fig. 4.9 (a) is a 118.0 (L) × 65.0 (W) × 151.0 (H) mm rail sample, on which three artificial cracks were induced. Fig. 4.9 (b) is a 164.0 (L) × 60.0 (W) × 128.0 (H) mm rail sample, on which two artificial squats were induced. Fig. 4.9 (c) shows a section of the rail in service at the Long Marston railway test track. The measured section is 210.0 (L) × 62.0 (W) mm and contains the joint of two plain rails and a shelling defect. Fig. 4.9 (d) shows a crossing nose measured at Long Marston. The measured section is 705.0 (L) × 98.5 (W) mm and contains the joint between the two plain tracks and the crossing nose at 285 mm from the vertex of the crossing nose. For clarity, all the defects and joints are numerated and pointed out with arrows. The quantitative information for each defect is summarised in Table 4.1.

<table>
<tr><td>(a)</td><td>(b)</td></tr>
<tr><td>(c)</td><td>(d)</td></tr>
</table>
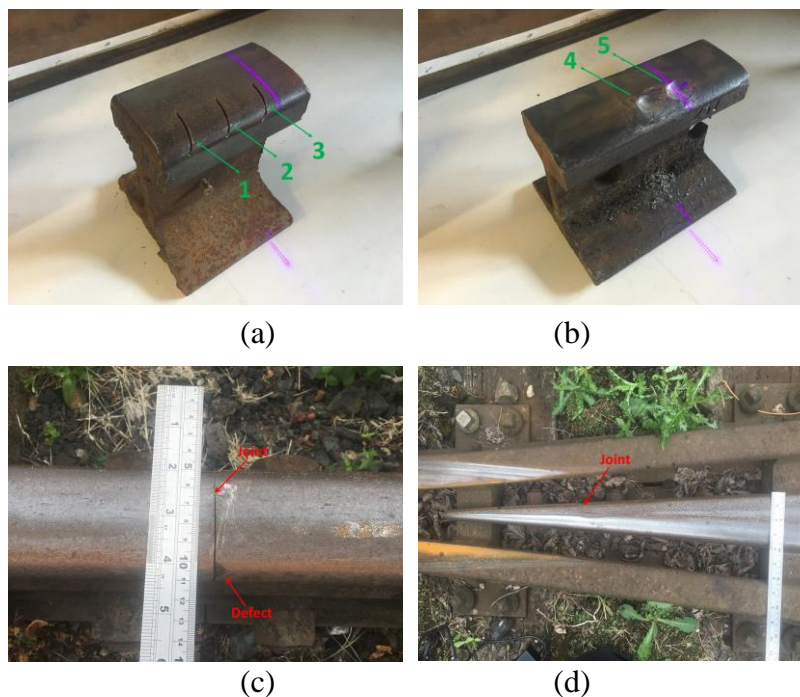
Fig. 4.9. Examples of laboratory and field tests

Table 4.1 Quantitative information of defects in Fig. 4.9

Defect	Surface length (mm)[a]	Maximum depth (mm)	Surface width (mm)[b]	Cross-section view (approx. geometry)
1	31.20	4.00	2.00	
2	28.30	3.00	2.00	
3	27.00	4.33	2.00	
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] Surface width refers to the size of the defect in the lateral direction of the rail.

Fig. 4.10 shows the polygon meshed laser measurement results for the samples in Fig. 4.9. Procedures and results before the polygon mesh follow the same route as the one in Fig. 4.4, and thus are not repeatedly presented here. It can be seen that the surface details of all the samples can be clearly captured. The size of each mesh element can be controlled by the density of interpolated points. In this case study, each mesh is about

(a)   (b)

(c)   (d)
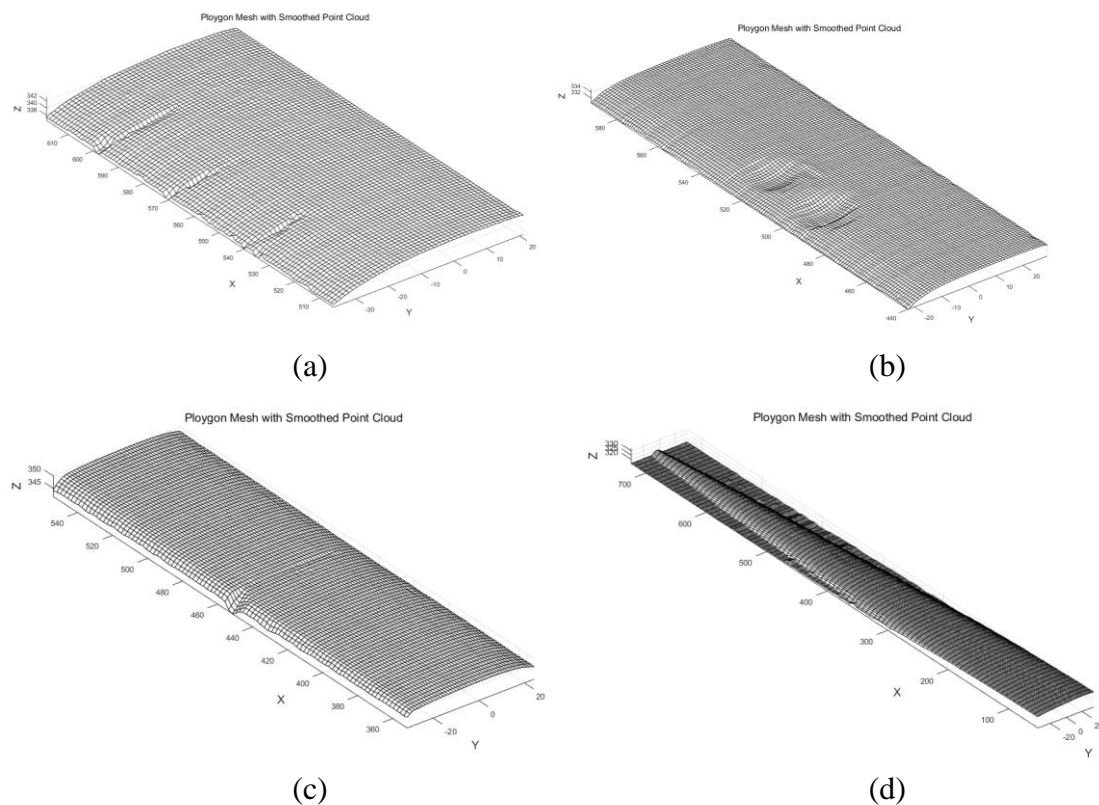
Fig. 4.10. Laser measurement results for the samples in Fig. 4.9

1 mm^2, which is small enough to represent detailed geometries of both the normal surface and the defect.

With 3D measurement results, more geometric information of the rail and the defect becomes available. This makes it possible to design more refined inspection strategies. As has been mentioned with the 2D measurement results, surface defects normally lead to sharp changes in depth or the orientation of normal vectors of data points. This is also applicable for 3D measurement results as the mesh element in 3D also comprises data points. Therefore, geometries related to the depth and the surface normal vector of mesh elements were analysed in this case study. Three geometric features were developed: Depth gradient, Face normal, and Face-normal gradient, each of which is in line with the detection of a specific type of surface defect. The definition of Depth gradient and Face normal-related detectors are the same as (4-3) and (4-4). Detailed mathematic derivations can be found in the published paper [10]. Fig. 4.11 shows the detection results for the proposed detectors with the defects shown in Fig. 4.10. It can be seen that the Depth gradient detector is more sensitive to defects like the cracks in Fig. 4.11 (a) and shelling in Fig. 4.11 (c) where the depth changes sharply, while the Face normal detector is more sensitive to detects like the squats in Fig. 4.11 (b) where the orientation of surface normal vectors is different to that of the normal surface. All the defects in this case study can be detected and characterised in 3D. Although there are some errors such as missing elements at the bottom of the squats in Fig. 4.11 (b), the 3D detection results are still considered to be more conducive to further defect evaluation than those in 2D.

(a) (b) (c)
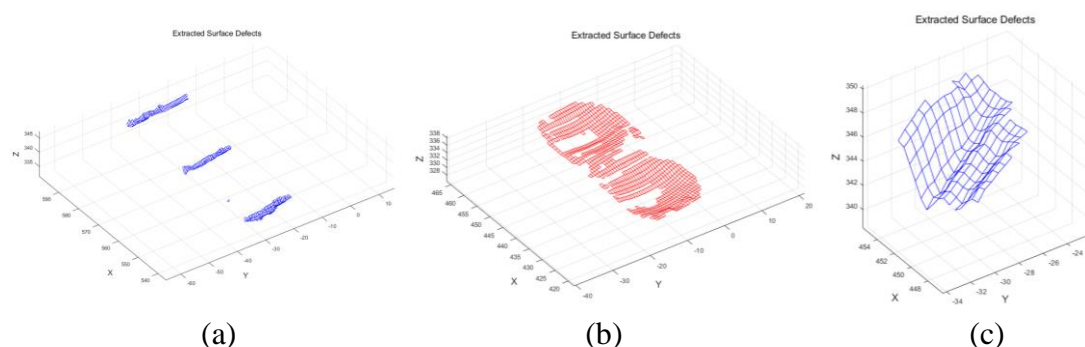
Fig. 4.11. Surface defect detection in 3D based on geometrical features

To further evaluate the performance of laser-based rail inspection and explore potential factors that could affect its application, the measurement accuracy was considered in

the form of relative errors between the laser measurement results and the actual specification of the object, calculated using (4-6):

$$\delta = 100\% \times \frac{|Actual\ value - Measurement|}{Actual\ value} \qquad (4\text{-}6)$$

It is worth noting that all measurement results are derived from the average value of repeated measurements (10 times in this case study) under consistent conditions to reduce the interference of human errors. Table 4.2 summarises the laser measurement results and the actual specification of rail samples in Fig. 4.9. It can be observed that the relative errors between the laser measurement results and the actual results are very small. These also contain random errors in the measurement process and errors caused by post-processing, e.g., the interpolation is a process of approximation in nature, and thus the laser sensor itself is considered to be accurate enough for rail measurement.

Table 4.2 Comparison of laser measurement results and the actual specification of rail samples

Rail sample	Length (mm)		Relative error (%)	Width (mm)		Relative error (%)
	Mea.	Actual		Mea.	Actual	
Fig. 4.9 (a)	116.0	118.0	1.69	63.5	65.0	2.31
Fig. 4.9 (b)	164.0	164.0	0.00	59.0	60.0	1.67
Fig. 4.9 (c)	211.5	210.0	0.71	61.0	62.0	1.61
Fig. 4.9 (d)	703.0	705.0	0.28	96.0	98.5	2.54

The measurement results for the rail samples are promising. Rail surface defects are another point of focus for this research. Accordingly, Table 4.3 compares the size of the extracted defects with their actual size. It is worth noting that rail joints are not considered to be defects, and thus are not listed here. It can be observed that most of the defects can be measured and characterised with a relative error under 10%, which demonstrates the accuracy of laser-based rail defect assessment. The defects that show large relative errors are typically the three surface cracks. After investigation, it was determined that this is mainly caused by the shadowing effects of the triangulation sensor. As shown in Fig. 4.12, shadowing effects refer to the shadowing of the laser line and the shadowing of the receiver [87]. *The shadowing of the laser line* is due to the fact that the fan-shaped laser line will disappear partially or completely behind steep edges.

Table 4.3 Comparison of laser measurement results and the actual specification of rail defects

efect	Length (mm)		Relative error (%)	Width (mm)		Relative error (%)	Depth (mm)		Relative error (%
	Mea.	Actual		Mea.	Actual		Mea.	Actual	
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**

(a)   (b)
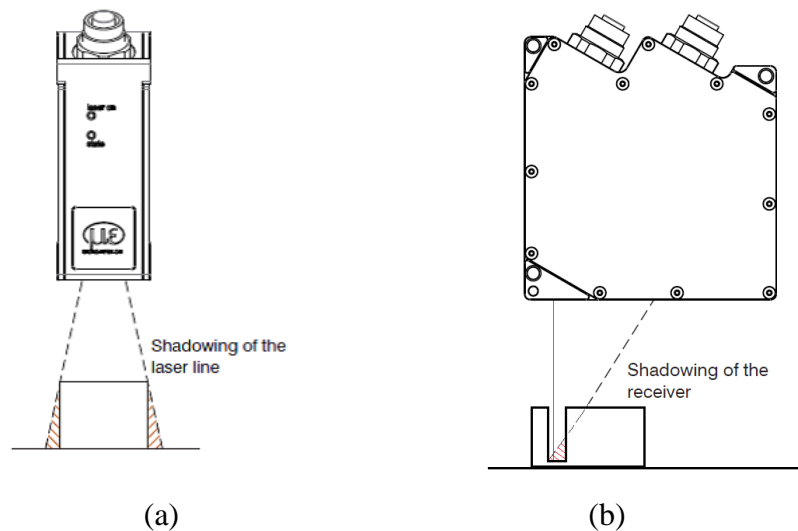
Fig. 4.12. Shadowing effects of the triangulation sensor

The areas behind the steep edges (marked with red hatched lines in Fig. 4.12 (a)) cannot be seen by the laser sensor. Therefore, surface defects such as the cracks in this case study that are partially distributed on the gauge corner area of the rail where steep edges are involved cannot be captured by the laser sensor. This results in a relatively large error in the length measurement result for the crack. *The shadowing effect of the receiver* is due to the fact that the triangular reflected laser line can be blocked in narrow gaps (Fig. 4.12 (b)). Therefore, defects having a large ratio of depth to width such as surface cracks can expose the shadowing of the receiver and lead to measurement errors in depth. The errors caused by the shadowing effects of the laser triangulation sensor are considered to be systematic errors. These can be minimised by standardising measurement procedures. For example, the shadowing of the laser line can be reduced

by pointing the laser sensor at different areas of the object and then combining the data from different viewpoints to generate a complete measurement result. The shadowing of the receiver can be avoided by pointing the laser in a direction where the reflected laser beam will not be blocked. Therefore, the laser sensor itself is considered to be reliable enough to measure surface defects in more cases.

Apart from systematic errors concerning the shadowing effects, there were also external factors encountered during the tests of this case study that may cause random errors. These random errors can influence the system effectiveness and thus are worth discussing here.

1) **Surface dirt and rust** are the common surface contaminations of the rail in the real-world environment. Such augmentations can affect the physical size of the rail, which in turn affect the accuracy of laser-based non-contact rail geometry measurement. Moreover, some tiny particles can interface with laser light by their reflectivity, which manifests as noisy points in the measured data.

2) **External light fluctuations** can affect most optical inspection instruments to some extent. For example, section 3.3 shows the importance of the stability of external lighting conditions for vision-based inspection systems. Compared to vision-based systems, laser sensors are more robust to external light changes as they actively project highly collimated laser light rather than relying on natural light reflections. This has been demonstrated in the field tests carried out under different lighting conditions in this case study [10]. However, external light fluctuations under complex light conditions may introduce reflections at the laser wavelength band. This can still interface with the laser light and result in inaccurate measurements.

3) **Mechanical vibrations** refer to the dynamic vibration of external structures during the movement of the laser. This can result in misalignment of the 2D rail profiles in 3D space, and thus was found to affect the quality of the 3D reconstructed model in this case study.

Considering these external factors, some additional measures or precautions may be necessary when using laser-based inspection in specific scenarios. For surface defect inspection, the ROI is the rail–wheel contact area where surface defects are mostly distributed. This area is normally dirt-free due to the rolling contact of wheels.

Therefore, surface dirt and rust have a minor impact on laser-based rail surface defect inspection. However, should accurate measurements be required for the area where surface contaminants are present (such as rail grease), pre-cleaning services may be expected before conducting a laser-based inspection. Additionally, conventional manual inspections using contact-based inspection tools and profile gauges may need to be carried out. In relation to external light reflections, most existing laser sensors operate at frequencies outside the usual visual light spectrum and have built-in interference filters or lenses for the suppression of external light. However, complex lighting conditions could still induce artificial lights that fall in the laser's frequency of operation, which is difficult to filter. Therefore, laser sensors should not be used in complex lighting conditions. Should they operate in such environments, additional measures such as a light shield may be needed to physically suppress external light. As a non-contact measuring instrument, the effect of mechanical vibrations is a common concern, which can affect the system's accuracy. The system introduced in this case study used two slide rails in the lateral direction to constrain the motion of the laser, which is sufficient for a static platform. However, large mechanical frames or vibration-damped mountings may be considered in other applications such as a mobile platform to decrease the amplitude of vibration and ensure the stability of the laser sensor during the measurement.

4.4 Summary

As one of the non-contact measuring instruments, laser triangulation sensors have their strengths when conducting surface geometry measurement tasks. This chapter identified the characteristics of laser-based railway track inspection techniques through a self-developed prototype that uses the structure of most existing laser-based rail inspection systems as a case study. Through numerous experiments in both the laboratory and the outdoor field environments, the main strengths of laser-based systems can be summarised as two points: 1) laser triangulation sensors can accurately measure rail surface geometries in most cases, benefiting from their imaging principle based on active laser light reflections, which shows more tolerance to the complex surface and environmental conditions than most vision-based systems that are based on passive light reflections; 2) 2D triangulation sensors can not only be used to measure 2D rail cross-section profiles but also are able to measure 3D surface geometries by using the linear nature of rail to align 2D profiles along the longitudinal direction of the

rail. Compared with 2D measurements, 3D measurements contain more complete geometric information of the rail with increased dimension and allow for more inspection strategies. In this case study, dedicated 3D data processing has been developed in the form of a 3D surface reconstruction pipeline and three geometric feature descriptors for rail surface defect detection. 3D measurements in combination with dedicated 3D data processing can characterise rail surface conditions and surface defects more comprehensively than conventional methods working in the 2D space alone.

However, laser triangulation sensors also have their limitations. Firstly, the shadowing effects of the laser sensors were analysed. The shadowing effect of the laser line can affect the measurement integrity of the sensor, which makes it difficult to get complete rail profiles from a single deployment. The shadowing effect of the receiver can cause errors when measuring some specific defects. Secondly, the inspection strategies are still limited compared to the numerous algorithms available for vision-based rail inspection. Although the feature descriptors developed in this case study have shown the potential of 3D model-based surface defect detection, they still rely heavily on handcrafted rules or features and thus have difficulty in dealing with the complicated requirements in practical inspection tasks. More intelligent and diverse inspection strategies are, therefore, expected to fully make use of the high-precision and dimensional laser measurement results. Apart from these limitations, there are also external factors that can affect the effectiveness of the laser sensor. Some of them are avoidable with potential measures as discussed, while others can bring further issues. For example, to ensure the stability of the laser sensor during measurement, existing systems mostly rely on an additional mechanical frame. This in turn increases the size and decreases the DoF of the system, which limits its capability to carry out more complex measurements. All of these limitations and external factors are worth considering in the development of any new laser-based inspection systems.

5 LASER-BASED MULTI-SENSING TECHNOLOGIES AND THEIR APPLICATIONS

The previous two chapters (3 & 4) investigated the characteristics of vision and laser sensing methods and their performance as individual rail monitoring and inspection techniques. Motivated by their strengths and shortcomings in actual applications, the prototype of a multi-sensing system that integrates these two techniques, to combine their strengths and eliminate their shortcomings, has been developed. This chapter introduces the proposed multi-sensing system from two perspectives: 1) system design and 2) multi-sensing data fusion. All these works were also carried out under the scope of the S-CODE project [11], which ensures they satisfy the industrial requirements and the hypotheses for next-generation multi-sensing rail inspection systems. Next, numerous hand-held tests were carried out. Three indices were considered to evaluate the performance of the proposed multi-sensing system: system accuracy, system stability, and system speed.

5.1 Overview

The case studies in Chapters 3 and 4 indicated the strengths and weaknesses of vision- and laser-based sensing techniques with two specific application scenarios. The vision-based system in Chapter 3 mainly focused on monitoring the status of the switch rail, while the laser-based system in Chapter 4 focused initially on the inspection of surface defects in the railhead surface. Many works of literature in this field (including the previous sections of this research) do not differentiate between condition monitoring systems and inspection systems since both of them typically measure some aspect of the object and its performance. However, to design a multi-sensing system, it is helpful to first discuss the differences between condition monitoring and inspection. This helps us to understand the focus of the new system and to ensure that the importance of the strengths and weaknesses observed from every single component can be accurately understood. In his thesis about railway S&C inspection, Marius Rusu made a good comparison of condition monitoring and inspection techniques [9]. This can be summarised as follows:

(1) Condition monitoring systems are generally employed to provide condition monitoring of assets. This indicates that the scope of condition monitoring is focused on developing problems with the aim of preventing impending failures before they impact normal operations. Condition monitoring systems normally rely

on the dependence between the fault and the effects they lead to rather than measuring the parameters that define the faults. For example, vibration and acoustic sensors are mostly used for the condition monitoring of roller bearings [91]. The sensors do not measure the fault itself which could be deformation or cracks in the bearing components. Instead, the fault is detected or predicted based on quantifying its extent and analysing its effect on the corresponding signals. Similarly, the case study in Chapter 3 introduced a system using camera modules to monitor the status of the switch rail. The toe gap and the edge of the switch rail were monitored by detecting the distinct texture features of related objects in the image. Therefore, the emphasis of condition monitoring is on continuous or frequent monitoring.

(2) Inspection systems, by comparison, are typically employed to measure and assess the properties that define the faults directly. For example, the laser-based system introduced in Chapter 4 measured the geometries of the rail and rail surface defects directly. Inspection is normally performed periodically based on specific requirements. The inspection strategy needs to conform to related standards to ensure whether the fault is within acceptable limits or tolerances at the time of inspection. Therefore, in comparison with condition monitoring, such inspection systems are naturally focused more on the accuracy of measurements.

Due to the nature of random distribution and the relatively small size of the target defects, the proposed multi-sensing system focuses more on its accuracy and stability. The case studies in Chapters 3 and 4 indicated that laser triangulation sensors are more accurate and stable than cameras in most cases since their imaging principle is based on active laser light reflections. Therefore, the multi-sensing system was designed to use the laser sensor as the primary measuring unit and the camera as the secondary module.

5.2 System design

The design of the new multi-sensing system initially aimed to eliminate the limitations of the laser sensor as the primary measuring unit, and thus to improve the overall performance of the system for rail inspection. Section 2.3.3 highlighted four limitations of laser-based rail inspection based on the review of existing systems: *(1) the integrity of rail profile measurement, (2) the size of the system, (3) the dimension of measurement, and (4) the diversity of inspection methods*. From the data point of view, the case study in Chapter 4 proved that the fusion of data from multiple sensors can naturally increase

the dimension of measurement and diversify the inspection methods. Therefore, the integrity of rail profile measurement and the size of the system are the main challenges that were considered in the system design stage.

As analysed in section 2.3.3, the integrity of measured profiles is mainly affected by the limited degrees of freedom of the laser sensor and the shadowing effect of the laser line. The bulky size of existing systems is mainly due to the dependence of the laser on the additional mechanical frame which, however, ensures the stability of the laser sensor in terms of its position and orientation relative to the rail. Therefore, the core of the system design is to increase the degrees of freedom of the laser sensor to eliminate the shadowing effect of the laser line and to overcome the dependence of the laser sensor on the additional mechanical frame to miniaturise the system, while also maintaining the high precision and high stability properties of laser sensors. The diagram in Fig. 5.1 illustrates the design concept of the proposed multi-sensing system. The laser sensor is designed to be released from the mechanical frame, which has full freedom of movement and rotation (Roll, Pitch, Yaw) in the X, Y, and Z axes. Such a system can be manoeuvred around the rail to provide measurements of the ROI without 'blind spots', and thus eliminates the shadowing effect of the laser line. The main contributions and methodology of such a design, which will be introduced in detail in the following subsections (5.2.1–5.3.2), can be summarised as follows:
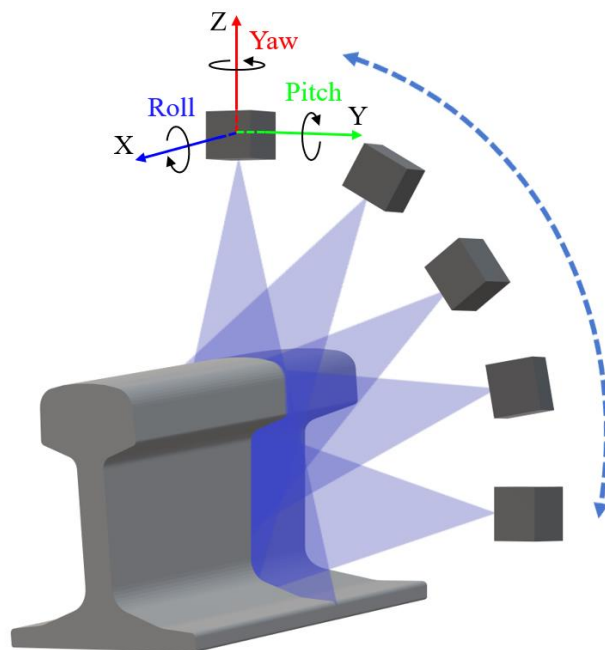
Fig. 5.1. Degrees of freedom in 3D space and measurement strategy of the proposed system

(1) To release the laser sensor from the mechanical frame, the position and the orientation of the laser sensor need to be tracked. For motion sensing, a simple and cost-effective method is the use of inertial measurement units (IMUs) as introduced in section 2.4. Therefore, an IMU module is integrated to track the motion of the laser sensor in 3D space. Additionally, vision-based object identification techniques are used in the form of a camera module to assist the progress of measurement by automatically evaluating the validity of the measured profiles. The integrated multiple sensors provide a more compact, autonomous, and portable prototype compared to most of the existing laser-based rail inspection systems.

(2) At the back end of the system, a multi-threaded data management strategy is proposed to fuse the data streams from the three different modules simultaneously. As discussed at the beginning of this section, the primary task of the multi-sensing system is to improve the integrity of rail profile measurement. This needs to accurately align or stitch the scans from different angles. With the multi-sensing data available, an optimised profile-to-profile registration method is developed to accurately align multiple scans and achieve greater integrity of rail profile measurement.

5.2.1 System architecture

According to the discussion above, the proposed multi-sensing system mainly consists of three modules: (1) a laser module, (2) a camera module, and (3) an IMU module. Before introducing the functionality of each module in detail, the diagram in Fig. 5.2 (a) illustrates the architecture of the whole system for clarity. The line arrows indicate the configuration between each module. Due to the large amount of data, the laser and camera modules are directly connected to the PC through standard Gigabit Ethernet and USB ports, respectively. The data from the IMU module are transferred through a 32-bit microcontroller development board using the IMU's SPI interface. Fig. 5.2 (b) shows the CAD model of the completed system. To obtain a compact prototype, all the components were fitted together to be able to fit into a customised housing. The connector provides the interface to external power supplies. The following sections (5.2.2 to 5.2.4) introduce the key components in detail.
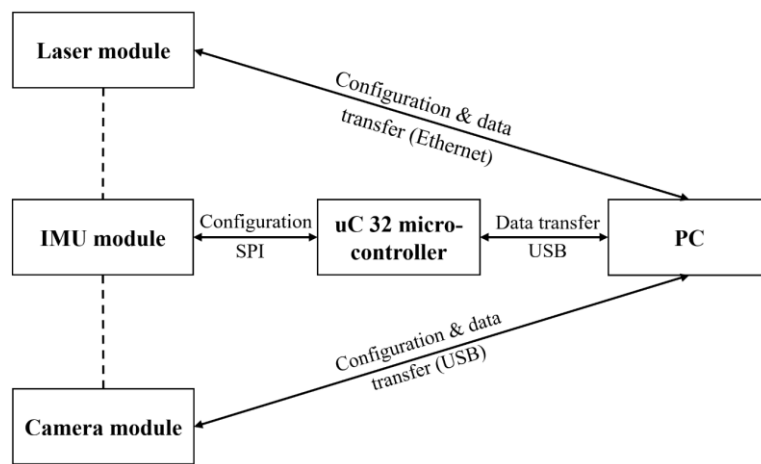
(a)
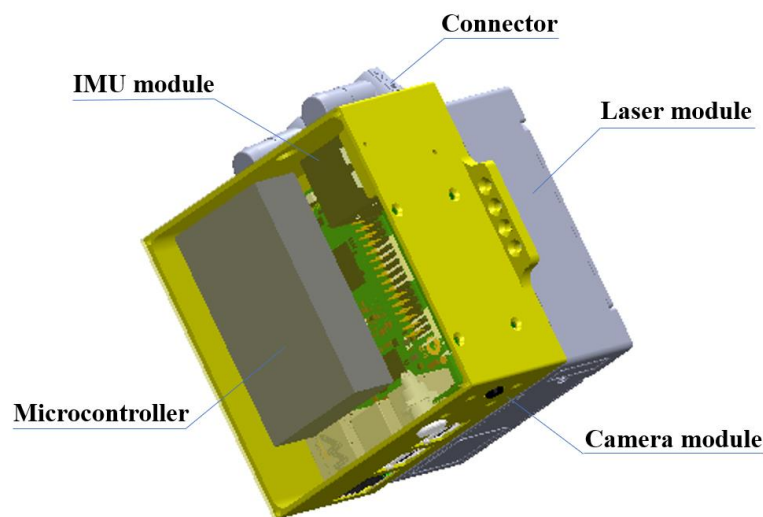
(b)

Fig. 5.2. Integrated multi-sensing system: (a) system architecture; (b) CAD model of the system

5.2.2 Laser module

The laser module is the primary measuring unit of the proposed system and thus the selection of an appropriate laser sensor determines the capability of the system. For industrial applications, laser sensors can normally be divided into two groups: blue laser scanners and red laser scanners. As shown in the visible light spectrum in Fig. 5.3, the colours here relate to the wavelength or frequency of the lighting source of the laser scanner. The wavelength of blue laser light is typically in the range of 400 to 475 nm, and the wavelength of red laser light is in the range from 650 to 780 nm. According to the Beer–Lambert law [92], a shorter wavelength of light source has less penetration

ability when facing the same medium. For the laser triangulation sensor used in this research, less penetration means more accurate measurement results. Therefore, a 405 nm blue laser triangulation sensor from Micro Epsilon was selected [26], which has a high-profile resolution for high-precision measurement tasks and a relatively compact size. Some key technical data of the laser sensor are summarised in Table 5.1.
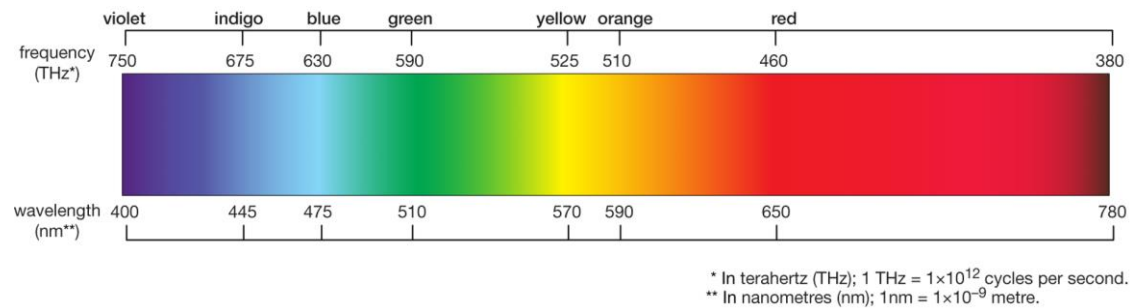
Fig. 5.3. Visible light spectrum: the correspondence between colours and wavelengths

Table 5.1 Technical specifications of the laser sensor

Model	scanCONTROL 2900-100/BL
Resolution in Y-axis	Max. 1280 points/profile
Profile frequency (standard)	Up to 300 Hz
Vertical measuring range (Z-axis)	125–390 mm
Horizontal midrange (X-axis)	100 mm
Spot size	Approx. 12 µm
Wavelength	Approx. 405 nm
Linearity	0.1% of the full-scale output
Operating temperature	0 °C to +45 °C

5.2.3 IMU module

The IMU is commonly used for motion sensing, which measures changes in rotation (known as Roll, Pitch, and Yaw) and the velocity and orientation of an object in the 3D space by combining a triaxial gyroscope, a triaxial accelerometer, and a triaxial magnetometer into one. The IMU was originally designed for the inertial navigation of transport such as aircraft, ships, and submarines. The recent development of miniaturised IMUs makes it possible to integrate them into more systems such as the

previously introduced SLAM system in the field of 3D modelling and the system in this research.

For the proposed system, the IMU module was used to track the motion of the laser sensor to allow its full freedom of measurement. Full freedom of the laser sensor for rail profile measurement has two meanings. First, it is the heading angle (α) of the laser sensor when it is manoeuvred around the rail (Fig. 5.4). Since the laser sensor itself does not know its position in 3D space, the measured profile will be inclined with an angle the same as the heading angle of the laser sensor at the time of measurement. And thus, an accurate measurement of α can help to initially rotate measured profiles back to normal so that they align with the vertical direction of the rail in the reference coordinate system. The second meaning is the relative angle (β) between the laser line and the longitudinal direction of the rail (Fig. 5.5). Since the laser sensor measures the profile of the rail, a valid measurement needs the laser line to be perpendicular to the rail length ($\beta = 90°$). Existing systems mainly rely on additional fixtures to access these angles, which in turn increase the system size and limit the degrees of freedom of the laser sensor. According to the requirements, an IMU (ADIS16405) from Analog Devices was selected, which outperforms in terms of sensitivity, accuracy, robustness, and reduced physical size among others [93]. It mainly contains a triaxial accelerometer
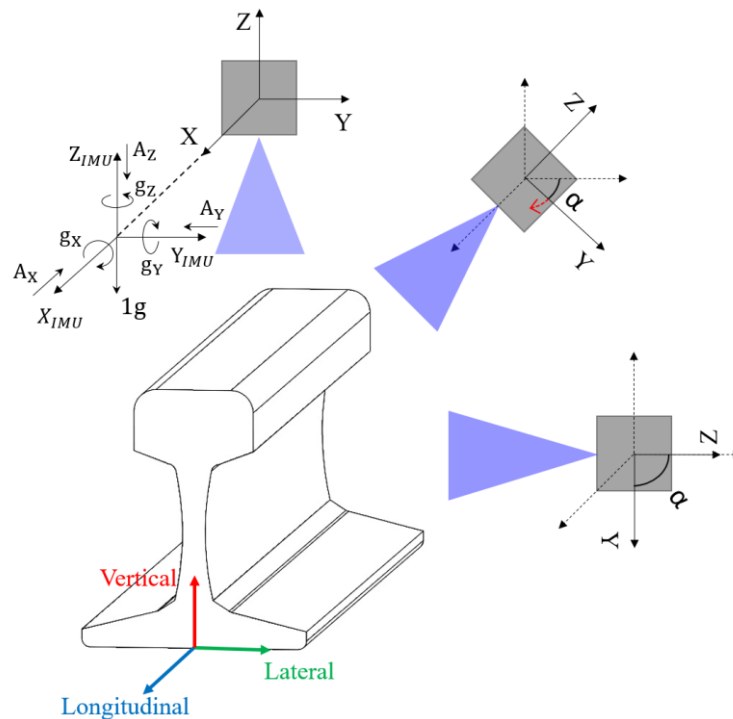
Fig. 5.4. Heading angles of the laser sensor

and a triaxial gyroscope, which can be used to calculate both accelerations $((A_Z, A_Y, A_Z)$, measuring range: $\pm 18\ g$, sensitivity: $3.33\ mg/LSB$) and angular rate $((g_X, g_Y, g_Z)$, measuring range: $\pm 300°/s$, sensitivity $0.05°/s/LSB$) with an output rate of up to 819 samples per second.
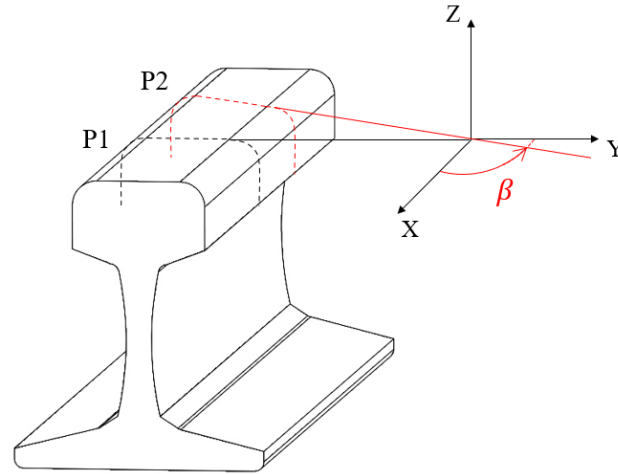
Fig. 5.5. Alignment angles of the laser line

For one measurement cycle, the laser sensor can have a heading angle (α) in two statuses: static and dynamic. Since the proposed profiling system was designed to have full freedom in 3D space, the laser sensor can start the measurement with a static heading angle (α_{static}) from 0° to 90° (relative to the horizontal plane). And then, the laser sensor can be manoeuvred around the rail with the heading angle ($\alpha_{dynamic}$) changing dynamically for a complete rail profile measurement. Both α_{static} and $\alpha_{dynamic}$ need to be tracked to match the corresponding profile during measurement. As shown in Fig. 5.4, when the laser sensor is stationary, the only stimulus to acceleration is gravity. This assumption helps to identify the horizontal plane for the laser sensor where the output acceleration is 0 g on both the X and Y axes, and 1 g on the Z-axis. Logically, any static heading angle (α) around the X-axis will create a projection vector of gravity on the Y-axis $A_{Y.Out}$, which can change from 0 to 1 g and conform to basic trigonometry [94]. Accordingly, α_{static} can be calculated using (5-1):

$$A_{Y.Out}[g] = 1g\ \times sin\ (\alpha_{static}) \tag{5-1}$$

The static heading angle can be calculated by Equation (5-1) simply. However, it cannot cope with calculating the dynamic heading angle as the movement and rotation of the system involve additional accelerations such as linear and centripetal accelerations.

These additional accelerations are difficult to distinguish from the gravity distributions, and thus can result in errors when calculating the dynamic heading angle. The gyroscope in the IMU module that measures the angular rate is able to solve this issue. By integrating the angular rate, the dynamic heading angle can be calculated more accurately. Therefore, the gyroscope is introduced (5-2) to calculate the dynamic heading angle.

$$\alpha_{dynamic} = \int_0^t g(t)dt \tag{5-2}$$

where $g(t)$ is the angular rate of the laser sensor, which could be $g_X(t), g_Y(t)$, or $g_Z(t)$, indicating the rotation relative to the corresponding axis.

It is worth noting that the IMU is not an error-free option for motion sensing. Especially for the calculation of the dynamic heading angle, accumulated error, also known as 'drift' can occur due to the constant measurement of changes and rounding off of calculated values. Therefore, the dynamic heading angle here is only used for the coarse rotation of the measured profile, and then it is corrected with the proposed profile-to-profile registration process. The detailed information will be introduced in section 5.3.

5.2.4 Camera module

The heading angle of the laser sensor itself can be tracked and calculated by the IMU. However, the relative angle (β) that is expected to be 90° is not accessible for the IMU since the laser sensor and the rail have their own reference coordinates. To identify this relative angle, visual augmentation is introduced with the integration of an HD camera (720 p/30 fps). Based on the discussion in Chapter 3, it is already known that the fundamental principle of a vision-based identification method is normally based on the distinct texture features of the object. As the laser line has higher luminance than the background, pixels in the image that correspond to the laser line appear with high contrast against the background. This makes it possible to identify the laser line accurately through image processing techniques. Accordingly, a blue laser line identification algorithm was developed for the camera module. Fig. 5.6 illustrates the key procedures of the blue laser line identification algorithm. Firstly, pixels corresponding to the blue laser line in each video frame are roughly extracted by subtracting the greyscale from the blue channel of the image. The laser line in the
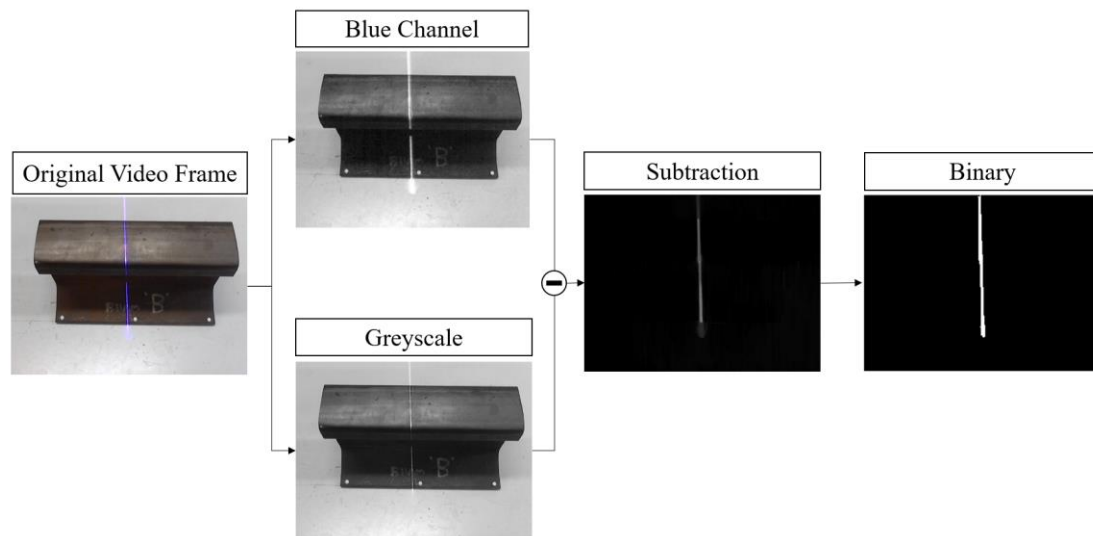
Fig. 5.6. Identification of the laser line through visual augmentation

subtraction can then be identified using Otsu's adaptive thresholding method to replace pixels with intensity values higher than the globally determined threshold with '1' and set all others to '0' [95], which yields a binary image indicating the position of the laser line.

During the experiment, it was found that two factors can affect the accuracy of the identification, for which two additional measures are taken to ensure more accurate laser line identification. Firstly, the shadowing effect of the laser sensor can result in discontinuities of the laser line at the vertical edges. These correspond to small gaps in the identification results as shown in Fig. 5.7 (a). To ensure the continuity of the detected laser line, a morphological 'closing' operator was applied to fill these gaps [96]. The operator works by conducting two basic processes: dilation and erosion, with

(a)   (b)

Fig. 5.7. Identification of the laser line through visual augmentation

the use of a structuring element of a specified size. Fig. 5.7 (b) shows the processed laser line, which ensures the continuity of the detected laser line and eases further processing. The second factor is the partial direct reflection of the laser beam from the rail surface. This can make the laser line appear wider in the image than it actually is, which could cause identification errors [97]. In this application, the light is projected vertically onto the rail surface; it was observed that the laser light diverges uniformly on both sides and the centre of the laser line in the image has the greatest intensity. Therefore, a morphological 'thinning' operator was used to erode the true ('1') pixels of the binary image, so that only the pixels near the centre are used to indicate the exact position of the laser line (Fig. 5.8 (a)). The laser line can then be identified and located accurately in the video frame (Fig. 5.8 (b)).

(a)   (b)
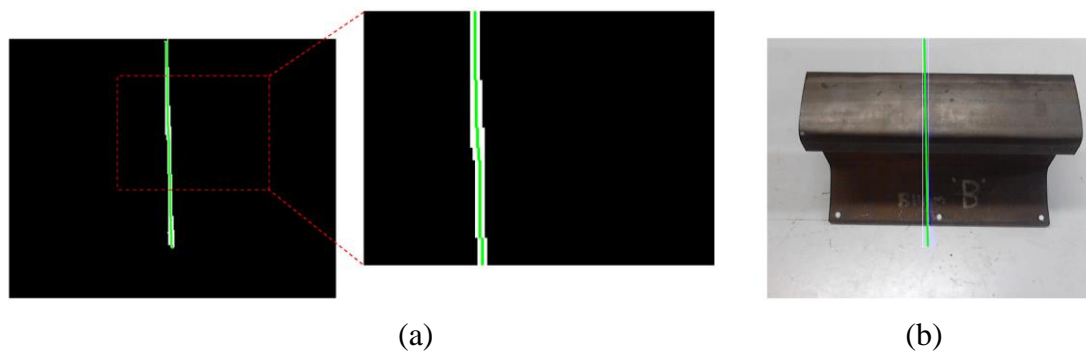
Fig. 5.8. Thinning operation of the identified laser line

With the laser line identified, the edge or the orientation of the rail needs to be identified to assess the relative angle (β) between the laser line and the rail. Commonly used edge detectors such as Sobel and Canny detectors were considered. It is known that these detectors work based on the sharp brightness changes of edge areas in an image. Therefore, surface cleanliness and a detectable contrast of the object against the background are normally necessary. However, these are elusive in the harsh working environment of the rail. As discussed in Chapter 3, the effectiveness of edge detectors can be significantly affected by factors such as surface contaminants and external light fluctuations. The inconsistent backgrounds that the proposed system would encounter can make the problem even worse. In comparison, the use of reference marker points has been proven to be able to make the detection more accurate and reliable in Chapter 3. Therefore, considering the required level of accuracy of the proposed system, a solution would be to identify some alternative components that align with the rail edge.

In this integrated system with visual augmentation, similar white reference markers (6 mm in diameter) with a black contour are used to indicate the edge of the rail. The Circular Hough transform is used in this application to detect the marker points based on their shape and brightness [98]. It should be noted that other feature descriptors such as Histogram of Oriented Gradient (HOG) could also be considered to detect the components such as rail clips in an actual application [99]. Fig. 5.9 (a) shows the visual augmentation results for the proposed system. The two vectors ($\bar{L}$ and $\bar{R}$ in Fig. 5.9 (b)) formed by the identified laser line and the marker points can then be used to represent the relative angle (β) simply as (5-3).

$$cos\,\beta = \frac{\bar{L} \cdot \bar{R}}{|\bar{L}| \cdot |\bar{R}|}$$

(5-3)

where $\bar{L}$ and $\bar{R}$ are the vectors formed by the points on identified laser line and the marker points, respectively.
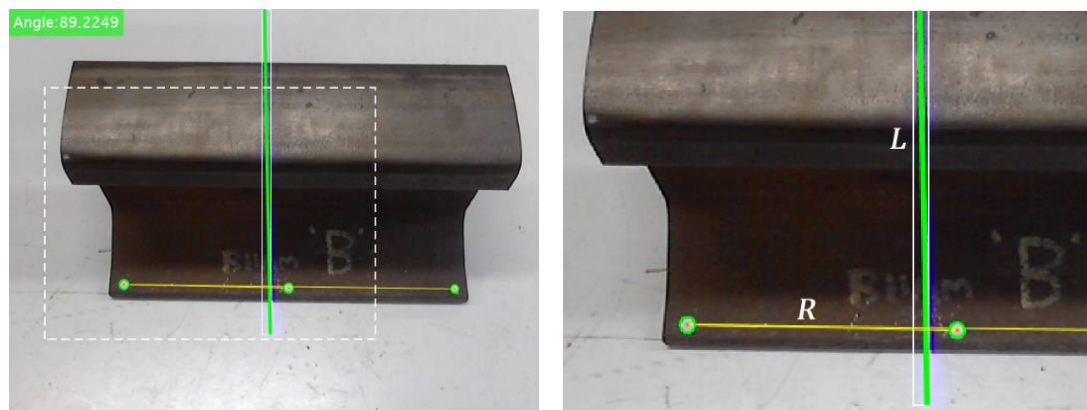
Fig. 5.9. Visual augmentation of the proposed system for the calculation of laser line alignment angles

5.2.5 Multi-sensing data fusion

The previous subsections introduce the architecture and the individual components of the proposed multi-sensing system. It is clear that the system includes three data streams from disparate sensors. To fuse them to generate more informative multi-sensing data, multi-threaded software was developed in Python to read, visualise, and save the data streams. Fig. 5.10 illustrates the basic logic of the software and the relationships between each thread. Since the camera module is used to evaluate the validity of the measured profiles, it works on the main thread. As vision-based angle measurement of a moving platform has its inherent inaccuracy ($\pm\,1°$ from experimental results), a

$\pm 2°$ tolerance is given to β. Once a valid angle (β) is identified, the system starts two further concurrent threads to read and visualise the data from the IMU and the laser. Reading and visualising data from three streams need a lot of computational resources and can cause delays. For example, the proposed system was connected to a PC configured with an Intel i7 series CPU with 16 GB of memory and an NVIDIA GeForce MX150 GPU with 2 GB of memory, which needs a delay of 50 ms between each frame. To read the data from the laser and the IMU simultaneously, a trade-off between the accuracy and speed of the system was made, with the sampling set to 20 Hz, and the resolution of the laser sensor set to 640 points per profile.
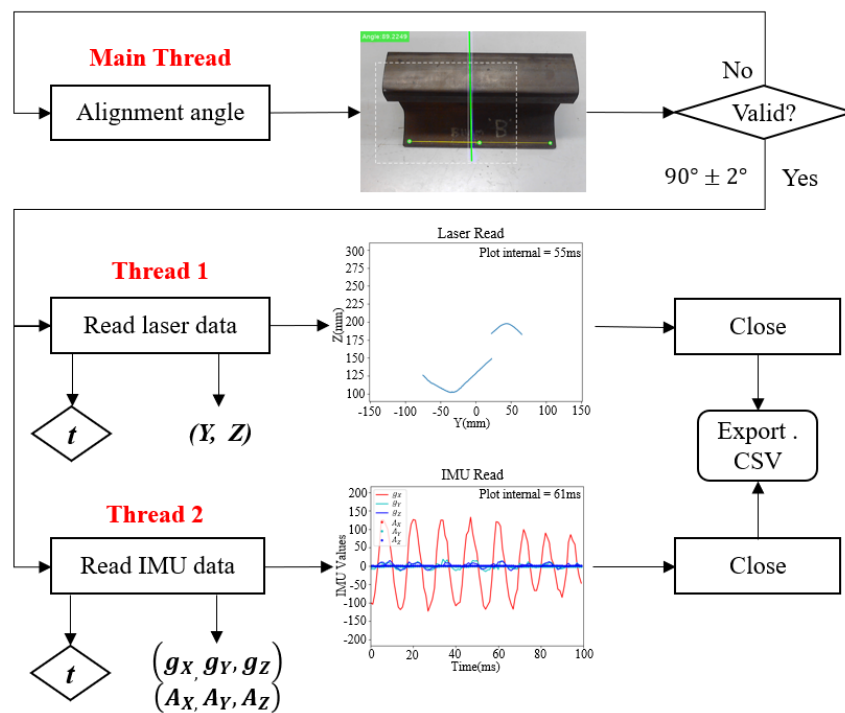
Fig. 5.10. Data streams and visualisation of the proposed multi-sensing system

The buffered multi-sensing data are saved into .CSV format. An example of the multi-sensing data set is shown in Table 5.2. Each profile contains 640 points (Y, Z), and its heading angles ($Roll, Pitch, Yaw$) and the accelerations (A_X, A_Y, A_Z). The dynamic heading angles are calculated from (5-2) using the corresponding angular rates (g_X, g_Y, g_Z) from the IMU. The accelerations are read directly from the IMU. With this information, the status and orientation of the laser sensor corresponding to each profile at the time of scan become accessible. For example, the dynamic heading angle of Profile 1 is 0 about all axes (Roll, Pitch Yaw). This means that the system starts the measurement from stationary. Meanwhile, the magnitude of gravity on the Z-axis (a_Z)

and its projection vector on the Y-axis (a_Y) indicate that the system has a static inclination angle with respect to the horizontal plane, which is $\sin^{-1}(0.68) \cong 42.84°$ based on (5-1). Similarly, Profile N has a dynamic heading angle of $20.09°$ about the X-axis relative to Profile 1. Therefore, the final heading angle of the laser sensor relative to the horizontal plane is about $62.93°$ ($42.84° + 20.09°$). This information will be used in the following procedure of profile reconstruction.

Table 5.2 Multi-sensing data fusion

Profile	Y (mm)	Z (mm)	Roll (°)	Pitch (°)	Yaw (°)	A_X (g)	A_Y (g)	A_Z (g)
1	60.80 ⋮ −75.46	206.85 ⋮ 123.76	0	0	0	0.01	0.68	−0.75
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
N	50.07 ⋮ −55.64	265.06 ⋮ 231.61	20.09	−1.18	−3.45	−0.04	0.98	−0.22

5.3 Complete profile reconstruction

The release of the laser sensor from an additional mechanical frame reduces the size and increases the degrees of freedom of the system. To improve the integrity of the measurement, the subsequent process would be to stitch scans from different angles together to reconstruct a complete rail profile model. Such a process is known as registration, which was proposed by Besl and McKay and has been mainly used for complete 3D modelling [68]. Generally, it includes two main steps:

(1) Use nearest-neighbour searching methods such as a k-dimensional tree (k-d tree) to identify the matched points between the models [71];

(2) Calculate the transformation matrix to minimise the Euclidean distance between corresponding points, using distance metrics such as the iterative closest point (ICP).

Since the simple process of registration was proposed, this method has also been used for shape equivalence comparison and has been migrated into rail inspection for comparison between the measured profile and the corresponding standard one. Fig. 5.11 illustrates the process of rail profile comparison using the ICP. Since the profile

measured by the laser triangulation sensor does not involve scale changes, the 2D transformation matrix (T) here only conducts an anticlockwise rotation $R(\theta)$ and translation offsets (T_y, T_z) on the Y and Z axes. It helps to register the measured profile (m) to the standard profile (s) so that the squared Euclidean distance metric as defined in (5-4) can meet the convergence criteria.

$$E(T) = min \sum ||s - Tm||^2 \qquad\qquad (5\text{-}4)$$
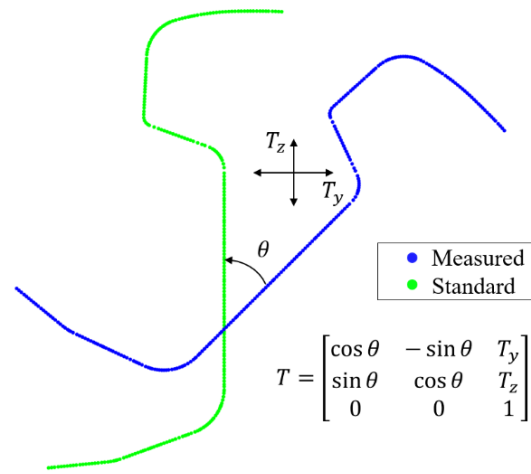
Fig. 5.11. Schematic diagram of rail profile alignment using ICP

5.3.1 Classic ICP registration

Similar to shape alignment, the strategy of the proposed profile reconstruction is also to find matched points between profiles captured from different angles and calculate the transformation matrix between them. However, the difference is that partial profiles captured from different angles have different shapes and hence less shape equivalence. Therefore, to accurately find matching points and create the transformation matrix (T) becomes more challenging. As an illustration, Fig. 5.12 shows two partial profiles captured from different angles and their registration result using classic ICP registration. It can be seen that the partial profiles are not registered correctly. The main reason for the poor registration result is that the object function (5-4) of the registration process is known as a non-convex function [100]. As an illustration, Fig. 5.13 shows a typical graph of a non-convex function. During the optimisation process, the expectation is to find the global minimum (green point), whereas a non-convex function may converge

(a)   (b)
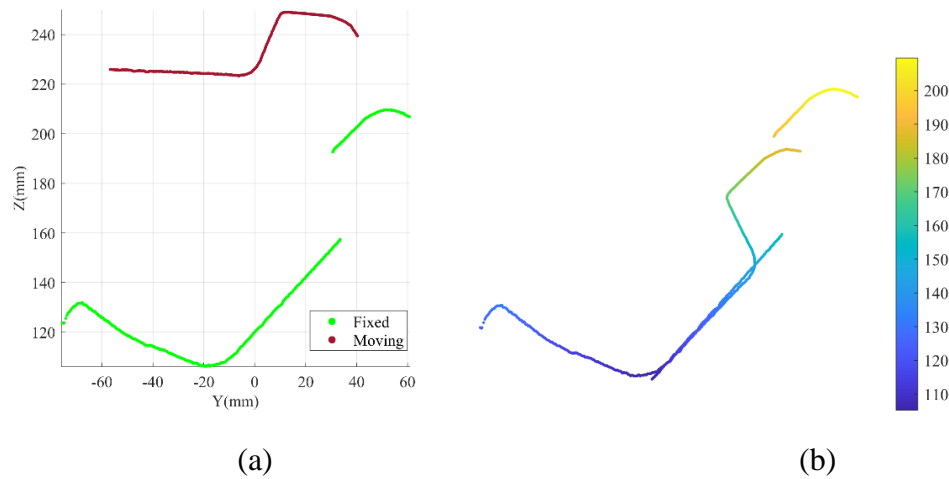
Fig. 5.12. Registration result of partial profiles: (a) raw sensor data; (b) registration result

to a local minimum (red point). The scattered initial position and the geometric differences of partial profiles increase their risk to converge to a local minimum. Existing solutions for this issue are based on either converting the object function to a convex one, for example by adding weight parameters to the function so that the global minimum can be achieved [63], or on obtaining a good initial position for the registration point sets to ensure convergence [73]. For the registration of a sequence of profiles captured from different angles, the pipelines associated with tuning the weighting parameters can be complicated and impose additional computational costs. Therefore, obtaining a good initial position is often considered to be a more practical solution. A good initial position for profile registration normally includes two factors: (1) the relative distance between profiles, and (2) the relative angle between profiles. For the proposed multi-sensing system, partial profiles are captured by rotating the system around the rail. Therefore, each profile has an inclination angle relative to the
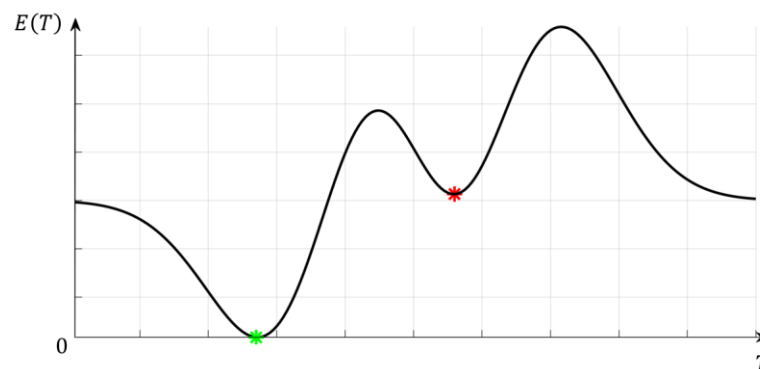
Fig. 5.13. Illustration of registration process based on non-convex object function

initial position of the system. This angle corresponds to the heading angle (α) of the scanner at the time of scanning. It is already known that the heading angles can be accessed from the IMU data, which makes it possible to create the corresponding inverse rotation matrix (R) using (5-5):

$$R = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \qquad (5\text{-}5)$$

This provides a coarse alignment for all the profiles before the registration procedure in (5-6):

$$m' = Rm \qquad (5\text{-}6)$$

where m is the original point data of the measured profile, and m' is the point data after inverse rotation. Fig. 5.14 (a) shows a good initial position of partial profiles after applying coarse rotation. The registration result in Fig. 5.14 (b) is significantly improved.

(a)   (b)
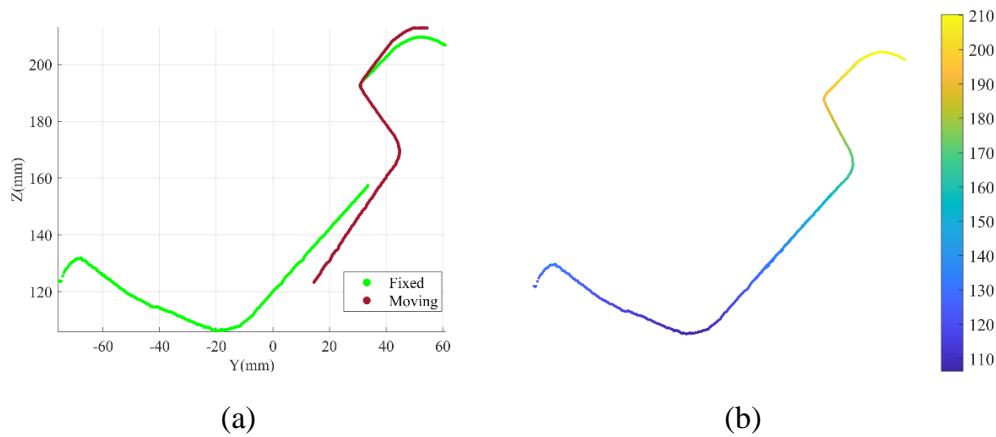
Fig. 5.14. Registration result of partial profiles: (a) sensor data after coarse rotation; (b) registration result

5.3.2 Optimised ICP registration with multi-sensing data

Fig. 5.15 shows the raw partial profiles of a complete measurement cycle using the proposed system. The aim of the registration process is to accurately stitch together these partial profiles using ICP registration to reconstruct a complete profile of the rail being measured. It can be seen that the profiles are spatially distributed, with which it is impossible to guarantee the quality of registration.
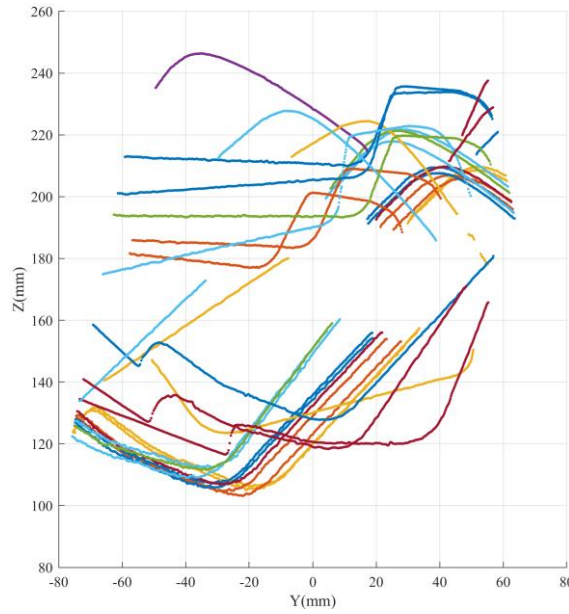
Fig. 5.15 Visualisation of profiles of a complete measurement

Building on the discussion in 5.3.1, an optimised ICP using multi-sensing data is proposed. Fig. 5.16 illustrates the process of the proposed method. To begin, the raw partial profiles are divided into two parts; the first profile is designated as the reference profile, establishing the reference coordinate system. The profiles that follow are set as 'moving profiles', which are registered to the reference profile to create the complete profile (Fig. 5.16 (a)). The reference profile has a static inclination angle. This is because the system starts the measurement with a heading angle (α_{static}) of around $50°$ (relative to the horizontal plane), which can provide adequate coverage of the rail. For the moving profiles, since the system is rotated around the X-axis during measurement, their dynamic inclination angles ($\alpha_{dynamic}$) are in the range from around $-40°$ to $+50°$ from the expected initial position. To give all the partial profiles a good initial position, an initial rotation matrix is created from the IMU data to apply an inverse rotation to each profile first to reduce the relative distance between profiles before the registration (Fig. 5.16 (b)). For the reference profile, the exact static inclination angle (α_{static}) can be calculated using (5-1) and the corresponding gravity vector distributions (A_Y). For the moving profiles, the corresponding dynamic inclination angles ($\alpha_{dynamic}$) can be approximated by (5-2) using the angular rate around the X-axis ($g_X(t)$).
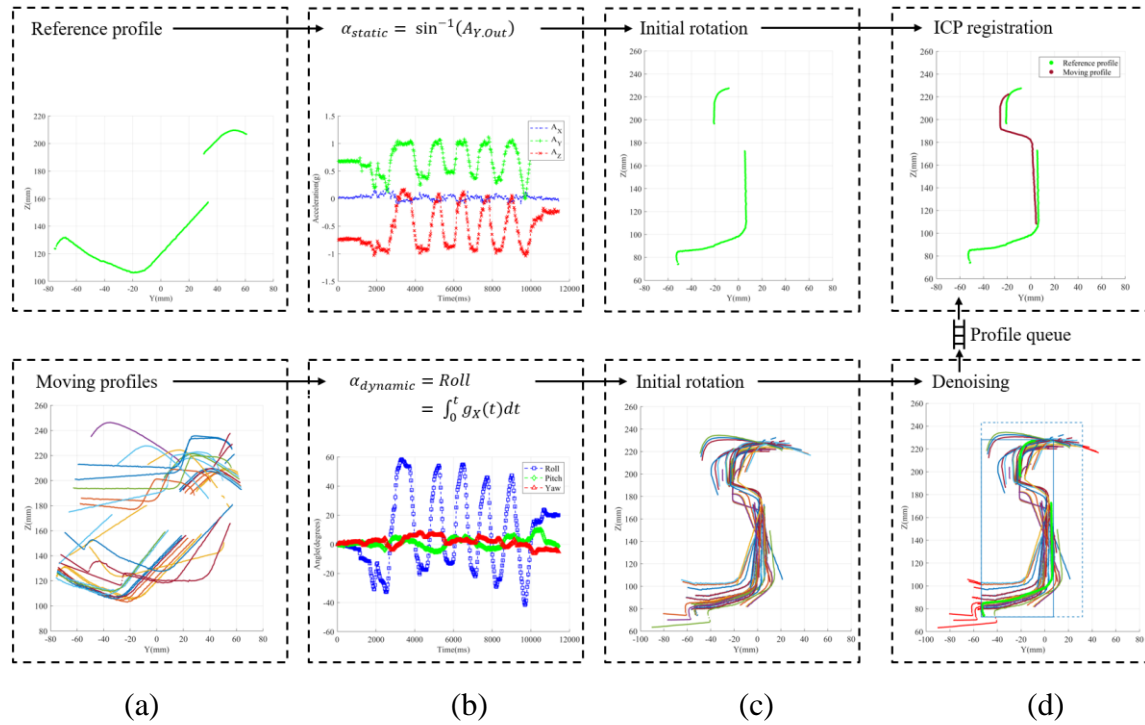
Fig. 5.16. Operation progress of the proposed profile reconstruction method:
(a) categorisation of raw partial profiles; (b) calculation of static and dynamic heading
angles; (c) results of initial rotation; (d) registration of profiles in sequence

When applying the inverse rotation, the axis of rotation is important. Equation (5-6) rotates all data points around the origin of the reference coordinate system. The profiles are measured, however, by rotating the system around the rail, and the centre of this rotation may not be the same as the origin of the reference coordinate system. This could result in profile dispersion around the origin after the initial rotation, increasing the relative distance between profiles even further. To address this issue, (5-6) can be modified to (5-7):

$$m' = R\big(m - Median(m)\big) + Median(m) \qquad (5\text{-}7)$$

The $Median$ approach in Equation (5-7) ensures that the origin of the coordinate system for each profile aligns with its mid-point before applying the inverse rotation. This procedure makes rotation based on the IMU data affect only the relative angle between profiles, not the distance between them. The reference profile and the moving profiles following the initial rotation are shown in Fig. 5.16 (c). The reference profile is rotated to correspond to the vertical direction of the rail, and the moving profiles are more centralised around the reference profile. The Euclidean distances between these

partial profiles are, therefore, significantly reduced. This helps the object function to converge faster to a desirable minimum solution with fewer iterations during registration.

To further increase the performance of the registration operation, a tailored denoising process (Fig. 5.16 (d)) is added after the initial transformation to eliminate redundant measurements at the ends of profiles (red portions outside the dashed box in Fig. 5.16(d)). Taking into account the symmetry of the rail, a rectangular box is built over the reference profile based on its height and breadth with a customisable tolerance. Any data points that fall outside the dashed box's designated region are eliminated.

Following the initial transformation and denoising steps, the transformation matrix can be calculated using ICP registration to reduce the distance between corresponding profiles and align moving profiles to the reference profile in a sequential manner. Because of the high sampling rate and slow movement rates of the system, adjacent profiles have nearly identical spatial distribution and shape. As a result, not every profile needs to be processed, and those processed can be aligned to their neighbouring profiles before combining the local transforms to ultimately align with the reference profile. The ICP in the proposed method is performed per n profiles, with n ranging from 3 to 10 depending on the geometric complexity of the object and the measurement speed of the system. This strategy can be illustrated as in Fig. 5.17 using a group of vectors to represent profiles, among which m_0 indicates the reference profile and others indicate the moving profiles. When n is equal to 3, the first ICP is applied on Profile 3

$$T1 = \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 & T_{y1} \\ \sin\theta_1 & \cos\theta_1 & T_{z1} \\ 0 & 0 & 1 \end{bmatrix}$$

$$T2 = \begin{bmatrix} \cos\theta_2 & -\sin\theta_2 & T_{y2} \\ \sin\theta_2 & \cos\theta_2 & T_{z2} \\ 0 & 0 & 1 \end{bmatrix}$$

$$T3 = \begin{bmatrix} \cos\theta_3 & -\sin\theta_3 & T_{y3} \\ \sin\theta_3 & \cos\theta_3 & T_{z3} \\ 0 & 0 & 1 \end{bmatrix}$$
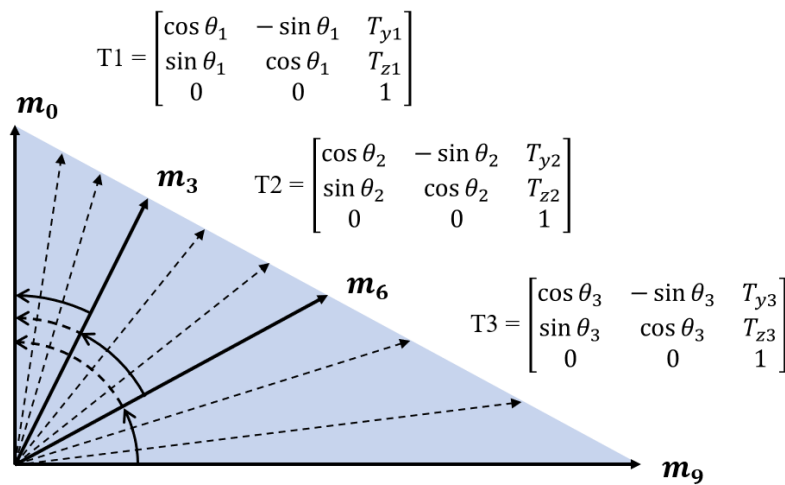
Fig. 5.17. Strategy for registration of a sequence of profiles

(m_3) calculating a transformation matrix T1. It transforms m_3 to m_0. The second ICP applied on m_6 generates the transformation matrix (T2) between m_6 and m_3. m_6 can then be transformed to align with m_0 by multiplying T1 and T2. This eliminates the need to repeat the computations for the route from m_3 to m_0. Similarly, a multiple of T1, T2, and T3 transforms m_9 to m_0. This process is repeated until all of the measured profiles have been transformed. Aside from lowering the computational cost of registration, a slight discrepancy between neighbouring profiles can enhance registration accuracy greatly by avoiding the local minimum.

Following the steps in Fig. 5.16, a reconstructed profile is shown in Fig. 5.18 (a). The railhead, web, and foot, as well as the transition areas between them, are successfully measured and reconstructed. Since the registration process aims to iteratively minimise the distance between profiles, the root-mean-square error (RMSE) identified as (5-8) is utilised as an assessment index.

$$RMSE = \frac{1}{M}\sum \sqrt{\left(m_i' - Tm_{(i+n)}'\right)^2} \tag{5-8}$$

where m_i' and $m_{(i+n)}'$ denote the neighbouring profiles to be aligned, T is the rigid transformation determined by the ICP, and M is the number of matching points. The average Euclidean distance between these matching points is returned by the RMSE.

To avoid any confusion between registration error and actual rail wear, a threshold is utilised. Any ICP with an RMSE greater than 0.05 mm is recognised as invalid, and the moving profile associated with it is discarded. This ensures that among all the candidates only the most accurate candidate measurements are maintained for the final reconstruction. The fitness curve in Fig. 5.18 (b) illustrates the registration process in this case; 16 moving profiles are registered through the optimised ICP, 13 of which are considered valid and aligned to the reference profile. From the very beginning of the registration, the average Euclidean distance between registered profiles typically ranges from 0 mm to 4 mm. During the registration process, most of these distances can converge to satisfy the criteria within 10 iterations. The final RMSE of each ICP operation is listed, giving an average registration error of less than 0.02 mm. These results demonstrate the performance of the proposed profile reconstruction method in terms of both accuracy and efficiency.

(a)

(b)
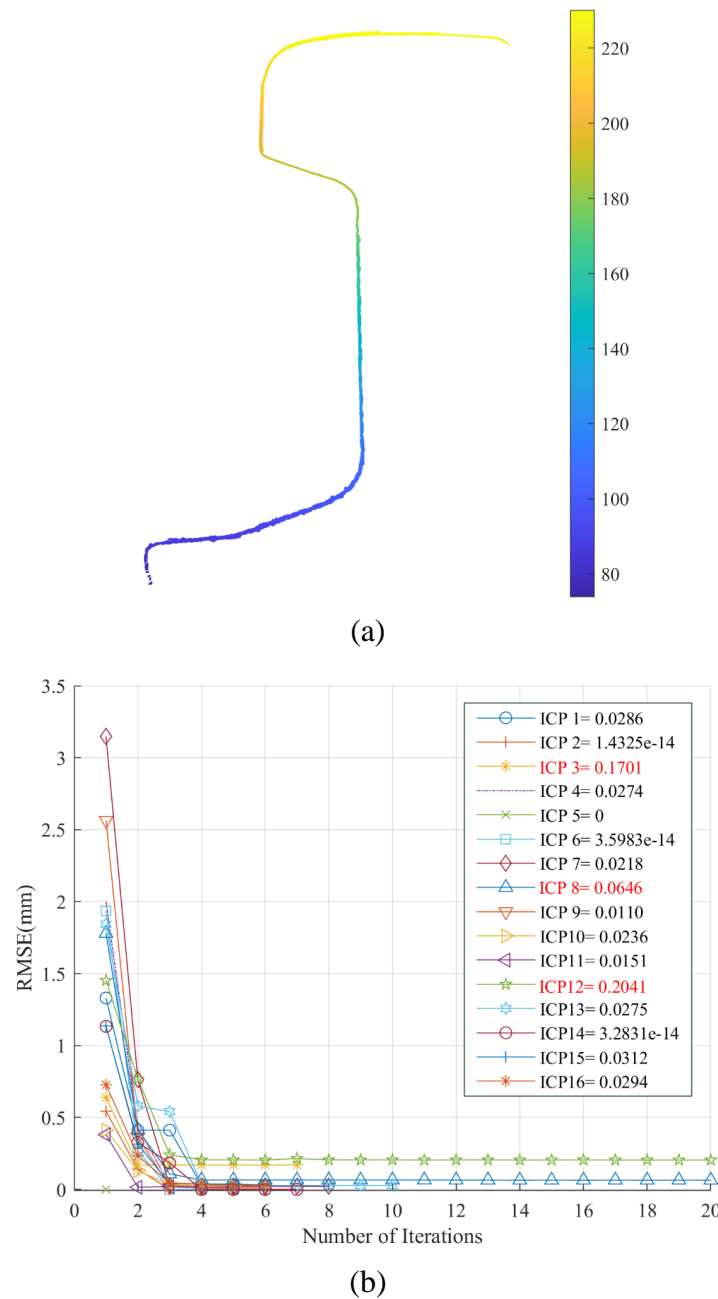
Fig. 5.18. Results of the proposed rail profile reconstruction:
(a) reconstructed profile; (b) fitness curves of the progress of reconstruction
with final RMSE of each ICP displayed

5.4 System tests and evaluation

The proposed system was tested by hand-held deployment with three indices considered: system accuracy, system stability, and system speed. Additionally, the performance of the system was also benchmarked by comparison with a commercial contact-based rail profile measurement system known as MiniProf. This section presents the test and evaluation results.

5.4.1 System accuracy

The system accuracy was evaluated in two ways: (1) comparison between the measured profile and the corresponding standard one; (2) comparison of measurements of neighbouring profiles. These are introduced in detail in the following two subsections.

(1) Comparison with standard profiles

To evaluate the accuracy of the system, the most common method is to perform a comparison between the measured profile and the corresponding standard one. Such a comparison can intuitively reflect the accuracy of the proposed system. Accordingly, different rail sections were selected. These included unworn rail sections and used rail sections to assess the effect of wear and rust. The rail sections used were flat-bottomed rail types (EN13674-1) with profiles 56E1 and 60E1, which are widely used in the UK [101]. Table 5.3 shows the standard dimensions of these rails for reference.

Table 5.3 Standard dimensions of the rail sections

Model	Rail height (mm)	Head width (mm)	Web thickness (mm)	Foot width (mm)
60E1	172.00	72.00	16.50	150.00
56E1	158.75	69.85	20.00	140.00

Fig. 5.19 shows the hand-held operation of the proposed system in the lab. All rail sections were measured under consistent conditions (e.g., comparable lighting and small ambient temperature changes ($+10\ °C$ to $+25\ °C$)), to avoid effects from these external factors. The measurement information is displayed on the PC simultaneously. This includes the identification result from the camera module, the real-time profiling result from the laser, and the real-time motion tracking result from the IMU. This visual information assists the user in pointing the laser at the rail with an appropriate alignment angle and stand-off distance while rotating the system around the rail. The rail profile is reconstructed from distributed partial profiles following the optimised ICP proposed in 5.3.2. The reconstructed profile can then be compared to the corresponding standard one using the four steps outlined below:

1) Down-sampling the reconstructed profile using a grid average method (0.50 mm). This helps to remove redundant data points as a result of the overlapping of partial
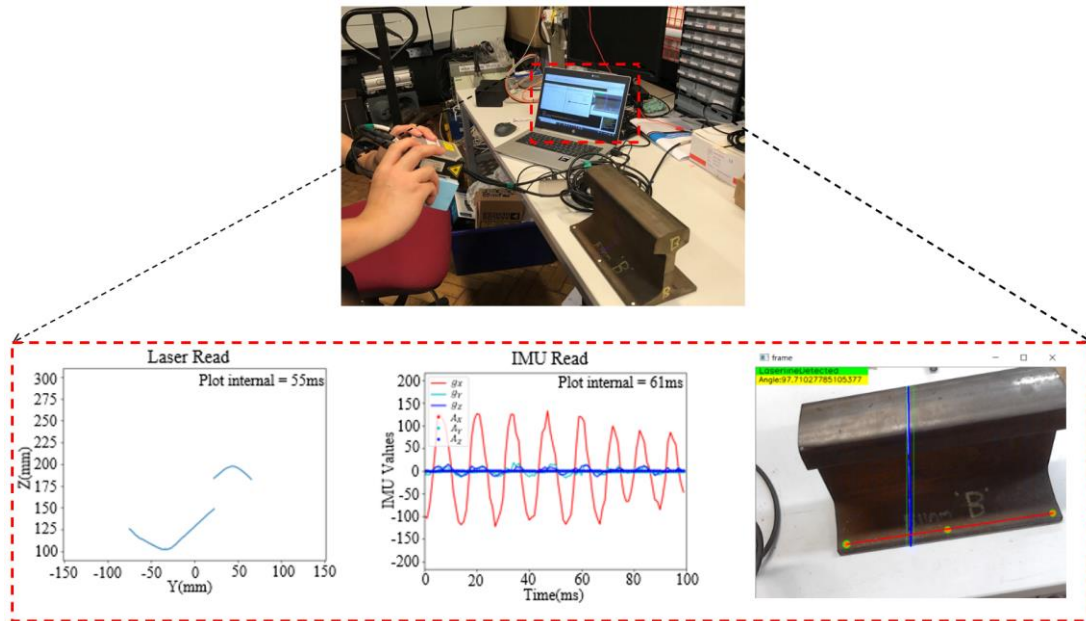
Fig. 5.19 Hand-held tests in the lab

profiles during the registration process, and thus reduce the computational complexity.

2) Extracting the standard profile with the same density as the down-sampled profile generated in step 1 from the corresponding standard CAD model.

3) Aligning the measured profiles to the standard one for shape equivalence comparison using the classic ICP introduced in 5.3.1.

4) Quantifying the degree of matching between the measured and standard profiles using RMSE.

For each rail sample category, Fig. 5.20 shows one measurement and the corresponding comparison results. The first row is the down-sampled measurement results, while the second row is the corresponding comparison results. The profile in Fig. 5.20 (a) is the measurement result for an unworn rail section with a 60E1 profile, which shows an outstanding match with the corresponding standard profile with an average RMSE of 0.07 mm. Fig. 5.20 (b) is the measurement result for a used rail section with a 56E1 profile. The actual height of the rail section used is 156.20 mm, which would correspond to a maximum abrasion loss of 2.55 mm around the railhead according to its standard dimension shown in Table 5.3. Based on the comparison result, the abrasion is accurately identified with a maximum loss of 2.50 mm. At the same time, the comparison result reveals a mismatch of approximately 1 mm around the rail foot area.

This can be caused by surface contaminants such as the build-up of dirt and rust on the rail foot due to long-term usage, which represents the real conditions of a rail in use as discussed in section 4.3. In this situation, the ability to measure the transition area between the railhead and the rail web of the proposed system plays its role in ensuring the measured profile can still be properly matched with the standard profile, to provide an accurate wear detection result.
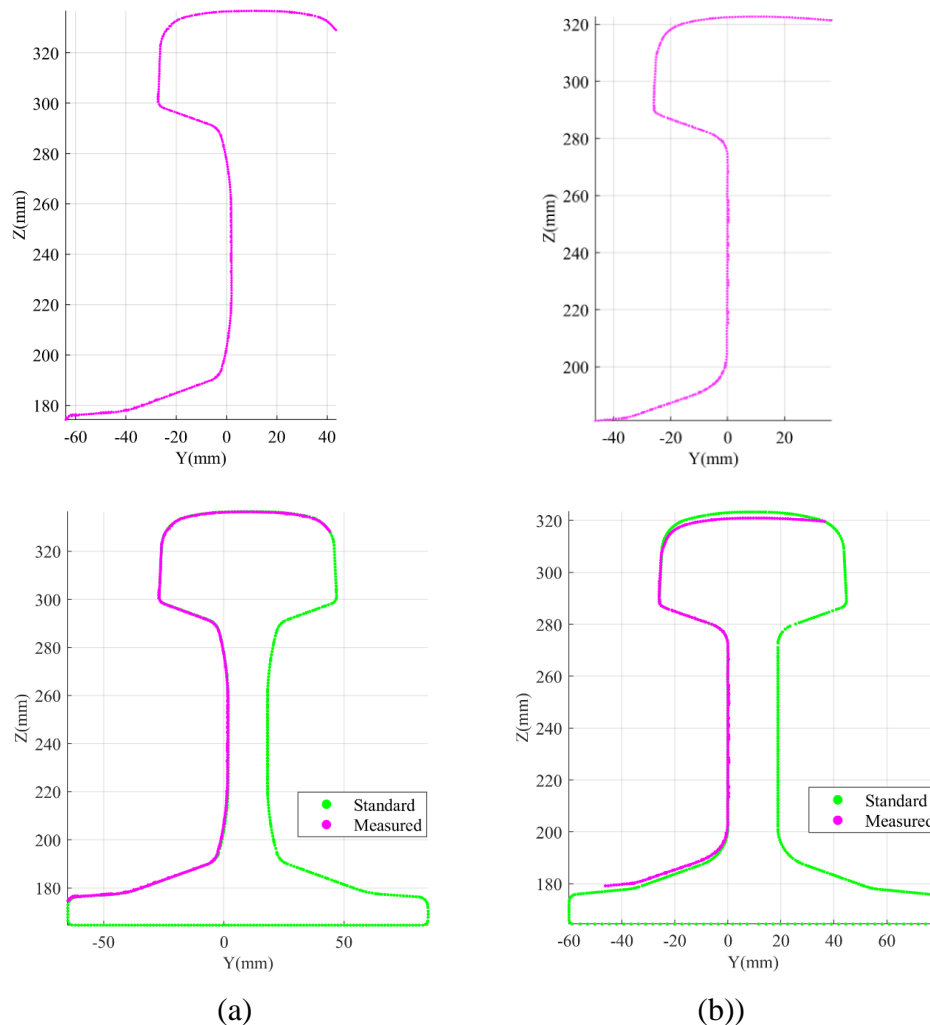
Fig. 5.20. Measurement and comparison results of the proposed profiling system:
(a) unworn 60E1 profile; (b) used 56E1 profile

(2) Comparison of neighbouring profiles

Currently, the comparison-based inspection method is reliant on the availability of the standard profile. However, the standard rail profile may not always be known, which brings difficulties for comparison-based wear assessment. The proposed system, therefore, allows the alignment of adjacent profiles along the longitudinal direction of the rail by using the unworn transition area between the underside of the railhead and

the top of the web, to identify regional wear without dependence on the standard profile. To verify the method, small patches with the specified thickness (1.50 and 0.50 mm, measured using a calliper with 0.01 mm resolution) were attached to the rail at the rail–wheel contact area to simulate abnormal profiles due to the wear. Fig. 5.21 (a) and (b) show the comparison results for adjacent profiles measured by the proposed system, and partially zoomed-in views of the ROI for clarity. The artificial wear can be identified without the use of the standard profile. The accurate alignment of adjacent profiles at the unworn transition area between the railhead and the rail web helps to identify the misalignment caused by the wear. In this example, patch thicknesses were taken from several misalignment points, as indicated by the arrow pairs in the partially zoomed-in views. The computed average thicknesses between these points are 1.61 mm and 0.56 mm, which are comparable with the actual thickness of the patches.

(a)

(b)
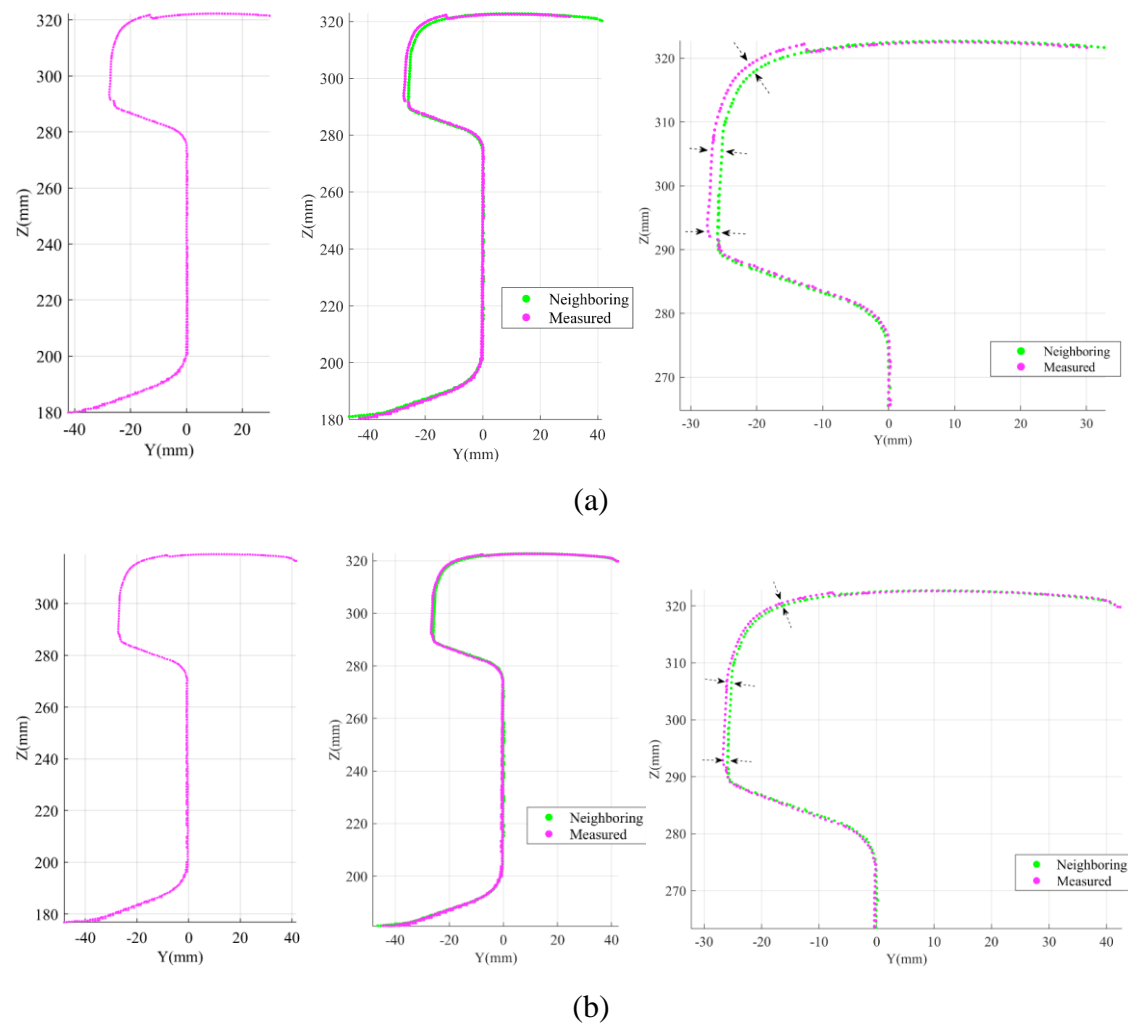
Fig. 5.21. Measurement and comparison results of the proposed profiling: (a) used 56E1 profile with 1.50 mm artificial wear; (b) used 56E1 profile with 0.50 mm artificial wear

5.4.2   System stability

For a measurement system, the system's stability is of equal importance to its accuracy. Stability refers to the capacity of a system to deliver the same measurement results over time when measuring the same sample. Mathematically, this can be quantified by the variance or the standard deviation of observed values. As discussed in section 4.3, random errors associated with external factors such as lighting conditions and mechanical vibration can also affect the effectiveness of the system. To accurately evaluate system stability, all measurements should be carried out under consistent external conditions. Therefore, an unworn rail section with a 60E1 profile was measured 10 times under the same conditions according to ISO 5725-1:1994. These include the same operator, the same illumination, and a short interval between profiles. As shown in Fig. 5.22, the purple line records the registration error of the 10 measurements, and the green line records the matching error between the measured profile and the standard profile. The registration error indicates how accurate the proposed registration process is by calculating the average RMSE of matching points between partial profiles that are registered together. The matching error indicates the shape equivalence between the reconstructed profile and the corresponding standard one by calculating the average RMSE of matching points between them. Overall, the average registration error is approximately 0.03 mm with a standard deviation of 3.5 μm. The average matching error is only 0.08 mm with a standard deviation of 7.2 μm. Both of these demonstrate the stability of the proposed multi-sensing system.
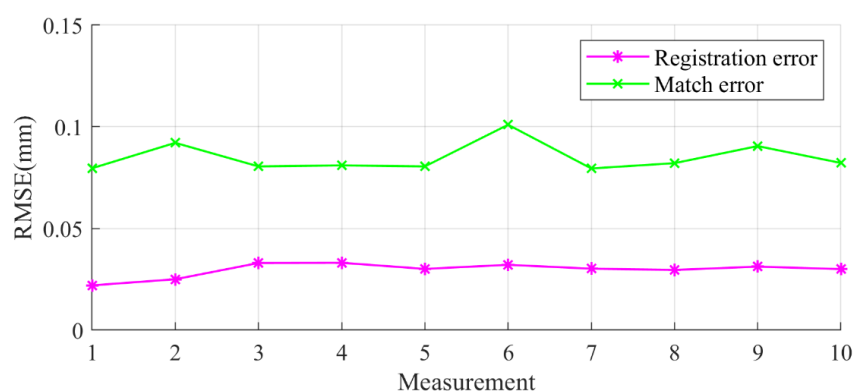
Fig. 5.22. System stability analysis

5.4.3   System speed

System speed determines the efficiency of a system when carrying out relevant inspection tasks. For the proposed system, the inspection process mainly includes three steps, namely measurement, profile reconstruction, and comparison. Accordingly, the system speed is analysed in terms of measurement time, profile reconstruction time, and comparison time. Measurement time here refers to the time required for a whole scan. The proposed system captures approximately 80 partial profiles for one measurement cycle, with a 50 ms time interval between each of them. Therefore, the measurement time is approximately 4 s. Of these 80 profiles, about 10 to 20 are identified as valid and can be used for the complete profile reconstruction, which can then be compared to the standard profile for rail inspection. The reconstruction and comparison time was analysed in two scenarios as shown in Table 5.4. The first scenario is with the real-time visualisation of the reconstruction progress enabled (as for the one shown in Fig. 5.16). The average reconstruction and comparison time are approximately 14.85 s and 2.15 s, respectively. The second scenario is with the real-time visualisation disabled. These times can be significantly reduced, to approximately 1.95 s and 0.23 s, respectively. It is worth noting that the profile reconstruction and comparison time may vary with different hardware and software configurations. The results here are based on the configurations as introduced in section 5.2.5, which can be used as an indicator for the efficiency of the proposed system.

Table 5.4 System operation time

Case	Measurement (s)	Reconstruction (s)	Comparison (s)
1	4	14.85	2.15
2	4	1.95	0.23

5.4.4   Comparison with contact method

For a non-contact profiling system, comparison with a full-contact profiling system is often of particular interest. Full-contact profiling systems take a direct measurement without any effect of intermediate mediums, and thus they are normally considered to be more accurate. Therefore, comparisons were also done with a commercial full-contact rail profiling system, known as MiniProf [102], to further evaluate the performance of the proposed system. The MiniProf was used to measure the same rail samples as the proposed system (Fig. 5.23).

Fig. 5.23. MiniProf full-contact rail profile measurement system

Fig. 5.24 shows the alignment of measurement results from the proposed system and the MiniProf for a standard profile. For clarity, enlarged views of the ROI are also included. The profiling results for the proposed system show good consistency with the results for the MiniProf, with an average matching error of less than 0.1 mm. As indicated by the dashed arrows in the enlarged view of Fig. 5.24 (a), the abrasion losses measured by the MiniProf and the proposed system are the same (2.50 mm at maximum). As shown in the enlarged views in Fig. 5.24 (b), in some edge areas that have concave discontinuities, the rolling probe of the contact system cannot track the sharpness of the discontinuity and tends to mismeasure edges as a slope associated with the roller diameter. In comparison, the proposed non-contact system can measure these edges more accurately. The ability to identify such discontinuities can be essential for the management or maintenance of specific defects, e.g., rail grinding. Table 5.5 lists other comparable metrics between the MiniProf and the proposed system. The results further demonstrate that the proposed system's performance is equivalent to that of the full-contact system while also providing greater coverage of the rail profile measurement.
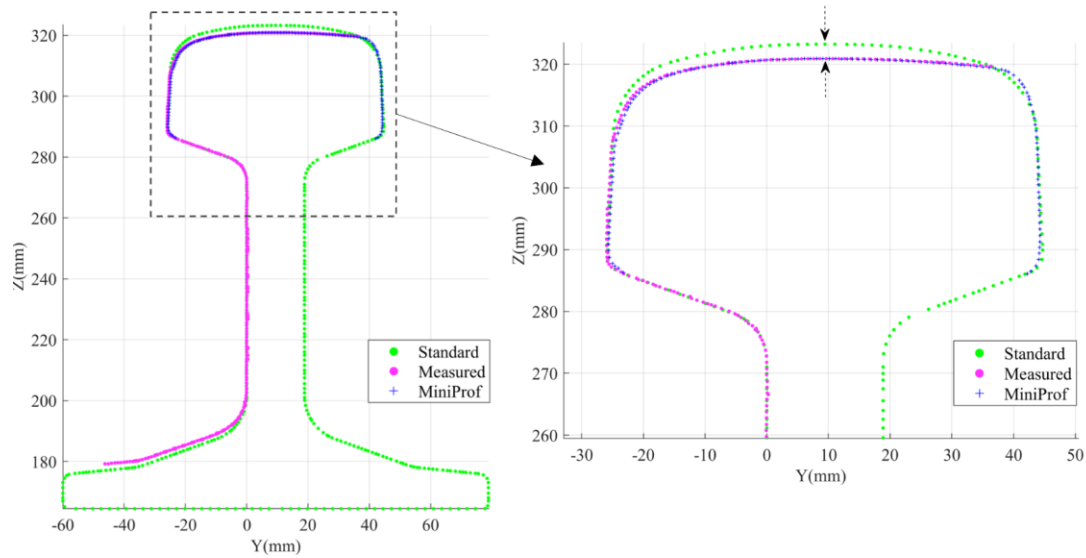
(a)
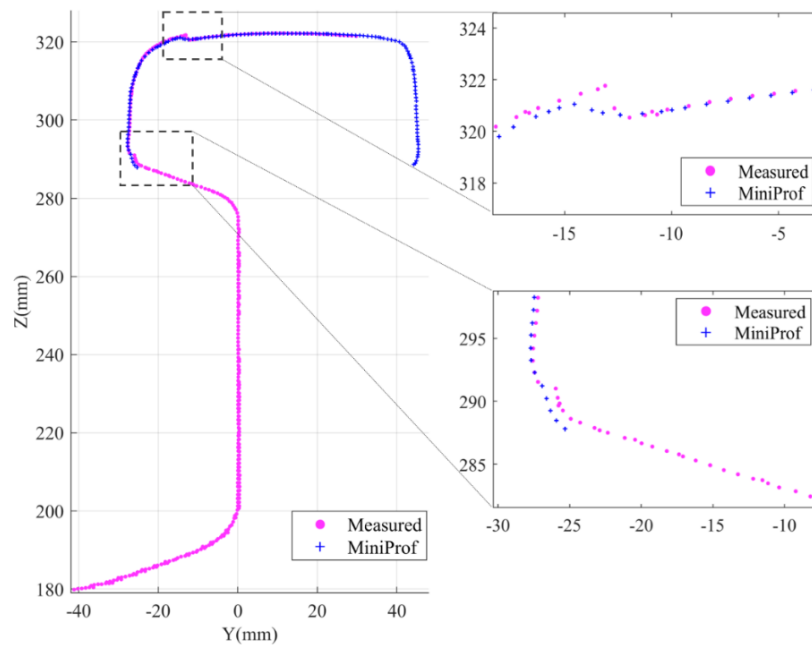
(b)

Fig. 5.24 Comparison with contact method: (a) measurement results of a used 56E1 profile; (b) measurement of discontinuities

Table 5.5 Comparisons with MiniProf

Metrics	MiniProf	Proposed system
Accuracy	$\leq 11\ \mu m$	$\leq 0.1\ mm$
Stability	$\pm 2.5\ \mu m$	$\pm 7.2\ \mu m$
Measurement speed	$\sim 5\ s$	$\sim 4\ s$
Coverage	*Rail head*	*Full profile*

5.5 Discussion about external factors

The main measuring unit of the proposed multi-sensing system is the laser sensor. Therefore, those external factors that have impacts on the laser sensor (discussed in section 4.3) may also affect the effectiveness of the proposed system. Meanwhile, the integration of new sensors can change the impacts of these external factors. Consequently, it is worth reviewing and discussing these factors further.

1) **Surface dirt and rust** are the common surface augmentations of the rail in use; they can affect the physical size of the rail, and thus result in inaccurate measurement results. Some surface augmentations with specific sizes may interface with laser light and cause noisy points in the measured profile.

2) **External light** is one of the major concerns of most optical instruments for accurate measurement results. Strong external light fluctuations can interface with the laser light by introducing the light in the laser's frequency of operation. In this circumstance, the measurement result can be inaccurate or even invalid.

3) **Mechanical vibrations** can affect the accuracy of the laser sensor in terms of profile misalignment. For the proposed multi-sensing system, mechanical vibrations can also affect the IMU module and the camera module, causing inaccurate motion tracking and laser line identification results. These can in turn affect the accuracy of profile reconstruction.

In practice, surface dirt and rust are typically present on the foot and web. For the proposed multi-sensing system, the primary ROI are the rail–wheel contact area and the transition area between the railhead and the rail web. The wheel–rail contact area is normally cleaned by the rolling contact with train wheels, while the transition area is likely to be dirt-free due to its overhanging nature. Therefore, the proposed system is found not to be greatly affected by surface augmentations. However, as mentioned in section 4.3, should accurate measurements be required for the area where surface augmentations exist, pre-cleaning services are expected before conducting a laser-based inspection. This observation is also valid for full-contact systems where significant contamination is present. Concerning external light, it is known that the laser sensor used in the proposed system has a built-in interference filter, which can suppress the external light to some extent. Meanwhile, it is anticipated that the user should avoid using the proposed system under unstable lighting conditions. However, should

complex lighting conditions occur, additional measures such as the use of a light shield or adjusting the exposure time of the laser sensor should still be considered. In relation to mechanical vibrations, since the proposed system is currently hand-held, such vibration of the unit is not significant. Should the system be deployed in other platforms, such as a vehicle or a drone, where mechanical vibrations occur, additional control measures such as a vibration-damped mounting or a stabiliser may be necessary to ensure the effectiveness of the system.

5.6 Summary

This chapter introduces the proposed multi-sensing system, which was designed based on the review of existing rail inspection systems in Chapter 2 and the case studies in Chapters 3 and 4. The design of the new system fully considered the characteristics of laser and vision sensors for rail inspection, which can be summarised in the following two points: (1) The system uses a laser triangulation sensor as the primary measuring unit and integrates an IMU module for motion tracking and a camera module to assist the progress of measurement. Such a design overcomes the dependence of the laser sensor on additional mechanical frames and yields a portable rail profiling system with increased degrees of freedom. (2) The multi-sensing system offers more dimensional and informatic multi-sensing data, with which an optimised profile-to-profile registration method was developed. The scans from different angles can be effectively registered. This achieves greater integrity of laser-based rail profile measurement.

The proposed multi-sensing system has been tested with hand-held deployment. The testing results demonstrate that the system can measure the rail profile with a controllable registration error of under 0.05 mm, and the complete rail profile improves the accuracy of profile matching for wear measurement, with an accuracy in the order of 0.1 mm on average. Comparison with a commercial full-contact profiling system further demonstrates the performance of the proposed system in terms of both the system accuracy, stability, and speed, and greater measurement coverage. Overall, it can be concluded that an effective combination of laser and vision sensors contributes to the improvement of their overall performance as rail inspection instruments. The two main contributions of the proposed multi-sensing system are: (1) the integrity of rail profile measurement is improved, enabling a more accurate and comprehensive wear assessment; (2) it demonstrates a new prototype for a laser-based multi-sensing rail inspection system with improved levels of automation and miniaturisation.

As a non-contact optical measurement system, those external factors that can affect the effectiveness of the system are discussed. Accordingly, additional measures that can help to ensure the system's effectiveness and broaden its range of application are analysed.

6 OPTIMISED INSPECTION APPROACHES WITH 3D AND MULTI-SENSING DATA

6.1 Overview

Chapter 5 introduced the proposed multi-sensor measuring system and demonstrated the strengths of the integration through hand-held deployment. The increased degrees of freedom and compact size of the proposed system broaden its range of application. Apart from the hand-held operation for complete 2D rail profile measurement, the multi-sensor system can also be deployed as an onboard measuring unit for more sophisticated inspection tasks. One that has been demonstrated in Chapter 4 would be 3D measurement and inspection. Laser-based 3D measurement can fully recover the surface geometry of an object, which allows greater potential for surface inspection.

For surface defect inspection, existing advanced strategies are normally based on computer vision algorithms using 2D images. However, vision-based inspection methods are susceptible to lighting and surface conditions as discussed in section 3.3. Laser-based inspection methods show more tolerance to the complex surface and environmental conditions; however, they are mostly limited to basic 2D geometric comparison or thresholding methods due to the lack of a substantial 3D dataset of laser-scanned surfaces and dedicated inspection methods. The proposed system has shown the ability to measure rail surface conditions accurately and generate a 3D rail surface efficiently. Logically, if related vision algorithms can be applied to process laser sensor data, the strengths of vision- and laser-based inspection methods can be combined, yielding some more advanced inspection approaches. This kind of complementarity is explored in this chapter by analysing the similarities between 3D laser sensor data and 2D images and introducing the process of translating between the two. Two new methods that draw from the concept of machine learning in computer vision are proposed: 1) end-to-end 3D pixel-level surface defect detection, and 2) 3D model-based surface defect classification and evaluation.

6.2 End-to-end 3D pixel-level surface defect detection

Computer vision-based surface defect inspection can normally be divided into two parts: defect detection and defect classification. Defect detection aims to report the presence and the corresponding regions of defects. As discussed in 2.2.2, existing vision-based defect detection methods can be categorised into three levels: 1) image-level, 2) regional-level, and 3) pixel-level. For rail surface defects, pixel-level detection is ideal

as it can not only locate defects but also describe their boundaries precisely. This is essential for further defect evaluation or management, e.g., rail grinding. Among the pixel-level defect detection methods, a deep learning model using fully convolutional networks (FCN) is one of the most effective methods as it achieves an end-to-end process that can automatically learn features and output pixel-level segmentation maps to indicate the corresponding class of each pixel with no need for any prior knowledge or handcrafted features [53]. However, since the detection is focused on each pixel of the image, pixel-level defect detection methods are naturally more susceptible to noise compared to the other two levels of detection. This limits their suitable areas of application, and thus existing pixel-level defect detection methods are mostly applied in scenarios where the lighting is favourable, such as the manufacturing lines of steel plates [57] and magnetic tiles [58]. In these environments, high-quality images that represent surface conditions are attainable. For the rail surface, many factors, as discussed in section 3.3, can distort the imaging process and result in low-quality images. Fig. 6.1 shows a 2D colour image of a defective rail sample captured by an HD camera in a laboratory environment. As can be observed, the quality of the images is distorted by the surface dirt and inconsistent surface textures. The regions that correspond to the real defects have a low contrast against the background (as indicated by the solid red box), and surface contaminations show a similar texture to those of real defects (as indicated by the dashed green box). All these factors can challenge the accuracy of pixel-level defect detection. To deal with these challenges, common solutions include increasing the size and diversity of the dataset, to ensure that useful features can be learned to identify real defects from others or to use additional measures such as image filtering and enhancement to ensure strong representation ability. These processes in return increase the cost in terms of both computational complexity and
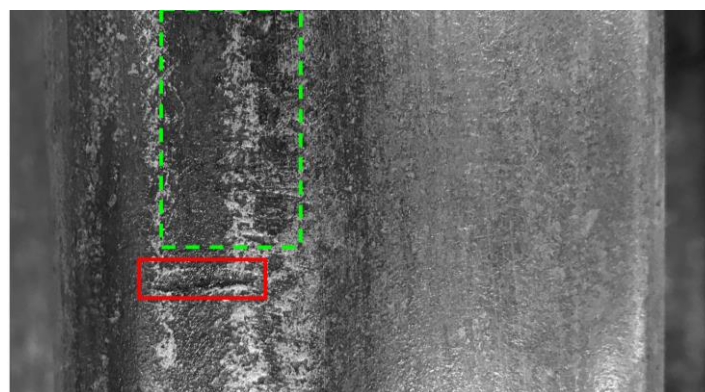
Fig. 6.1. Rail surface characterisation using an HD camera

data sources and make pixel-level defect detection intractable for most industrial applications where the practicability related to the cost must be considered. As a result, improving the imaging quality for an object with complex surface conditions such as the rail has been identified as a key factor in applying such pixel-level defect detection methods in more challenging areas.

As discussed in Chapters 4 and 5, unlike cameras that capture surface texture images using passive light reflections, laser sensors actively project structured laser light onto the object surface and receive reflected laser light to measure surface geometries. This enables them to directly measure geometrical changes induced by surface defects rather than depending on the correlation between surface defects and specific texture changes as cameras do. Therefore, laser sensors normally show more tolerance to complex surface conditions because general texture changes that can affect camera imaging quality have no impact on surface geometries. The case study in Chapter 4 has demonstrated that 3D point cloud data containing complete surface geometric information can be produced by aligning 2D laser profiling data along the longitudinal direction of the rail. As a brief recap, Fig. 6.2 shows the surface measurement results for the same sample as in Fig. 6.1 using the 3D measurement method introduced in (section 4.2.2-(2)). In comparison to the image in Fig. 6.1, the sequence of profiles measured by the laser sensor can represent the surface geometries accurately. The defects show high 'contrast' against the background with corresponding data points out of the normal surface (as indicated by the X–Z view of the region in the red box). Logically, pixel-level rail surface defect detection will become more capable if 2D images can be substituted by such high-quality laser sensor data. Inspired by this, an end-to-end pixel-level defect detection approach using an optimised FCN to process
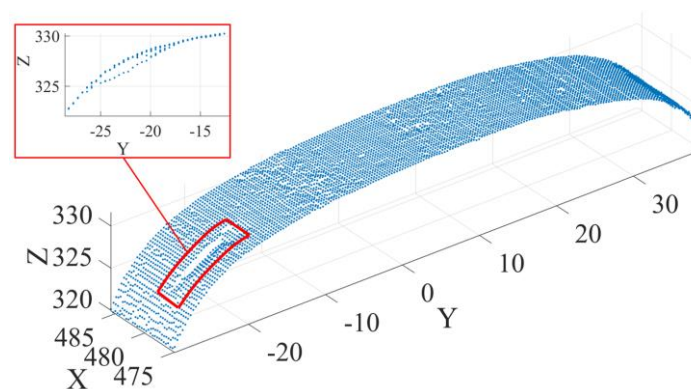
Fig. 6.2. Rail surface characterisation using the proposed 3D perceptual system

laser sensor data is proposed. The main contributions of this part of the work, which will be introduced in detail in the following subsections (6.2.1–6.2.5), can be summarised as follows.

(1) From a surface measurement perspective, an alternative solution by using the proposed multi-sensing system as an onboard unit has been deployed, to accurately characterise rail surface conditions.

(2) In terms of defect detection, transferability between 2D images and 3D laser sensor data was explored. A modulation strategy was developed to transform laser sensor data into depth-coloured maps which preserve geometrical information of rail surface and have similar properties as images. An optimised FCN was designed and trained successfully with the modulated laser sensor data, achieving end-to-end pixel-level rail surface defect detection.

(3) Compared to conventional pixel-level defect detection on images, the proposed pixel-level detection on laser sensor data enables a full characterisation of the defect, which provides not only the location and boundary of a defect but also recovers its geometry in 3D.

6.2.1 Surface measurement system

Fig. 6.3 shows the final deployment of the surface measurement system (SMS) by using the proposed multi-sensing system as an onboard unit. The structure of the platform has been introduced in section 4.2.1, so it is unnecessary to go into detail repeatedly. One significant improvement worth mentioning is that the camera module in the proposed multi-sensing system allows for automatic laser line alignment identification using the same laser line identification method as introduced in section 5.2.4. As the camera module works at a fixed point, the rail edge can be identified easily by using Canny edge detection and Hough transform methods [103], which provides an alternative solution for the rail edge identification in place of using marker points. Compared to the original system, this eliminates the error from manual operations and ensures that the laser line is perpendicular to the rail edge before measuring so that the profiles can be accurately aligned in the longitudinal direction of the rail to generate a coherent rail surface. It is worth noting that the motion of the laser sensor in this laboratory prototype is tracked by a displacement sensor. Therefore, the IMU module was not triggered.

(a)   (b)
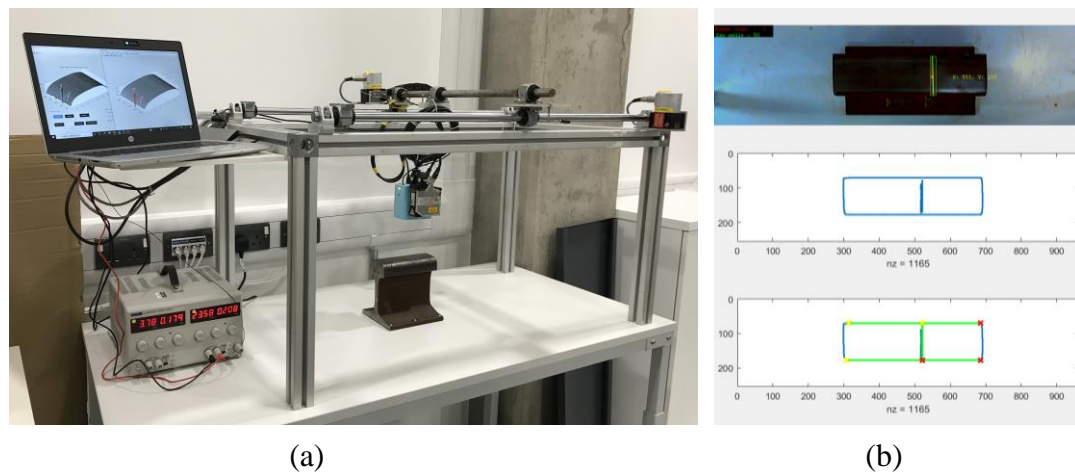
Fig. 6.3. Surface measurement system: (a) completed system in the lab; (b) vision augmentations.

6.2.2   Design of the FCN

It is already known that one of the most effective methods to achieve pixel-level defect detection is to use an FCN [53]. Existing networks use 2D images as inputs and output 2D segmentation maps. This research aims at achieving end-to-end pixel-level defect detection on 3D laser measurement data of the rail surface. Therefore, the inputs are 3D point clouds (the sequence of 2D profiles) from the SMS and the output should also be 3D segmentation maps. Accordingly, an optimised FCN was designed to achieve such a 3D segmentation task. Fig. 6.4 shows the architecture of the network. It consists of three modules: 1) a mapping module for 3D inputs, 2) a segmentation module (encoder and decoder), and 3) a mapping module for 3D outputs. The arrows represent the different operations as indicated in the bottom block of the diagram. As the learning process is to output feature maps through a learned linear transformation, each coloured box denotes the output tensor (or the feature map for the segmentation module) of the corresponding operator. The number of channels of the tensor is indicated on the top of the box. The x–y size is given at the lower-left edge of the box. An example of the input and the output of the network is visualised at the top of the network to facilitate understanding. A brief introduction of the three modules follows, details of which will be included in sub-section (1) to (3).

- First, an interpolation layer and a mapping layer constitute the mapping module. It takes 3D laser sensor data as the input and outputs modulated laser sensor data as 2D single-channel depth images. Each modulated 2D image preserves

the geometry of the associated 3D input while it also has high quality without the effect from surface textures and is fit for the segmentation module.

- Next, an FCN is used as a segmentation module. The segmentation module can be divided into two parts: an encoder and a decoder, according to the symmetric convolutional operations. It takes 2D single-channel depth images from the mapping module and corresponding ground-truth annotations as inputs for training and outputs 2D segmentation maps. The segmentation map denotes each pixel to a predefined class, either defective ('1') or normal ('0'), to achieve pixel-level defect detection.

- Finally, an output mapping layer symmetrical to the input mapping layer is added to map 2D segmentation maps back to the corresponding 3D laser sensor data, on which defects can be effectively reconstructed and characterised in 3D to support any subsequent defect evaluation process.
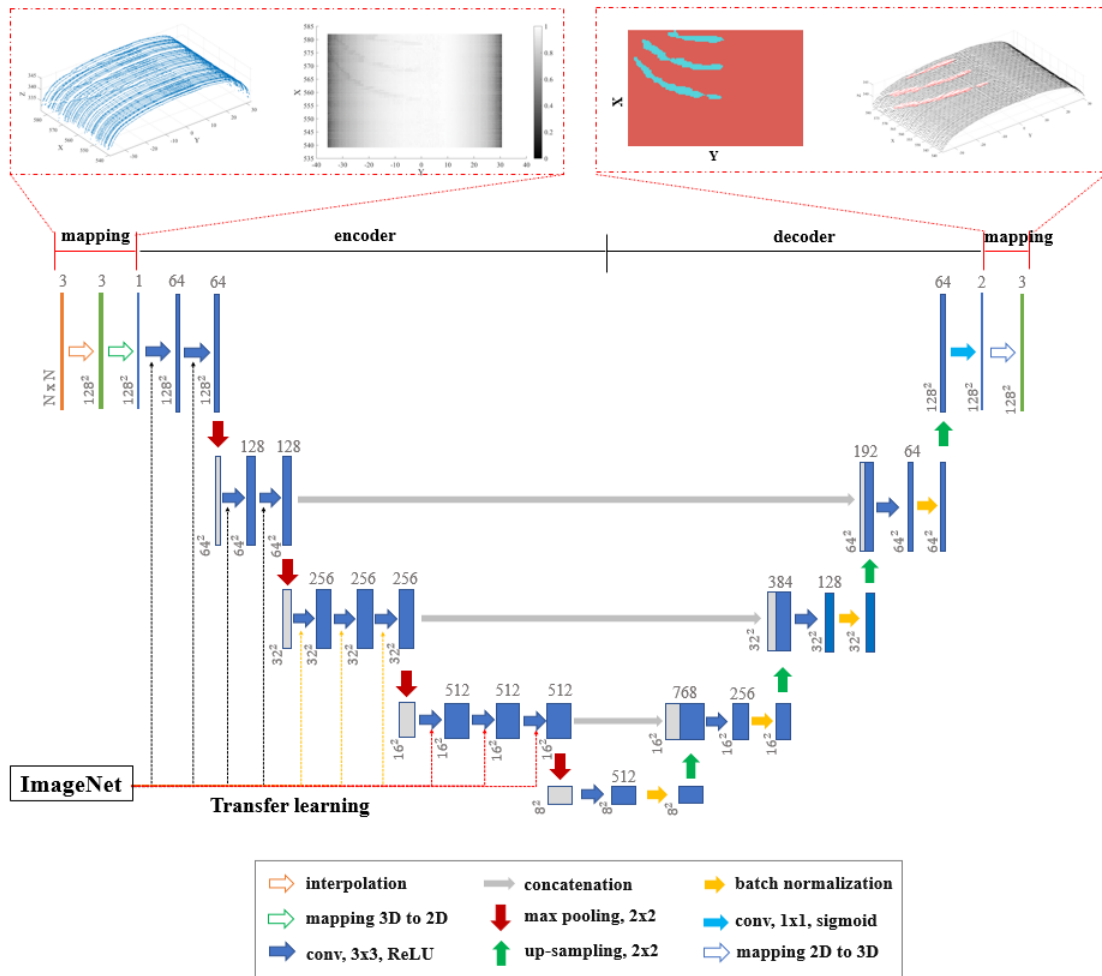
Fig. 6.4 Architecture of the proposed pixel-level defect detection network

(1) Mapping module for 3D inputs

Conventionally, a segmentation network takes 2D images and their ground-truth annotations as inputs. Each 2D image (I) is normally represented as a 3D tensor (6-1) when it is fed into the network.

$$I = (W, H, C) \tag{6-1}$$

where W and H denote the image's width and height, respectively, and C is the number of colour channels. For colour images, C is equal to 3, representing three colour channels (e.g., (R, G, B)). For greyscale images, C is equal to 1 as there is only a single colour channel.

As discussed at the beginning of section 6.2, the complex surface conditions of the rail can pose challenges to the suitability of pixel-level defect detection using 2D images. The method presented here uses 3D laser sensor data (as shown in Fig. 6.5 (a)) to replace 2D images. Similar to (6-1), the laser sensor data (P) can also be represented as a 3D tensor as (6-2):

$$P = (X, Y, Z) \tag{6-2}$$

where X, Y, and Z are the coordinates of a data point in the 3D space. When the 3D laser sensor data are viewed from the top as illustrated in Fig. 6.5 (a), X and Y can also represent the width and the height of the measured area. This is comparable to the W and H of a 2D image as indicated in (6-1). Therefore, the essential difference between the tensor of a 2D image and the 3D surface measurement data in this research is that in the tensor representing images, the third dimension (C) corresponds to the colour
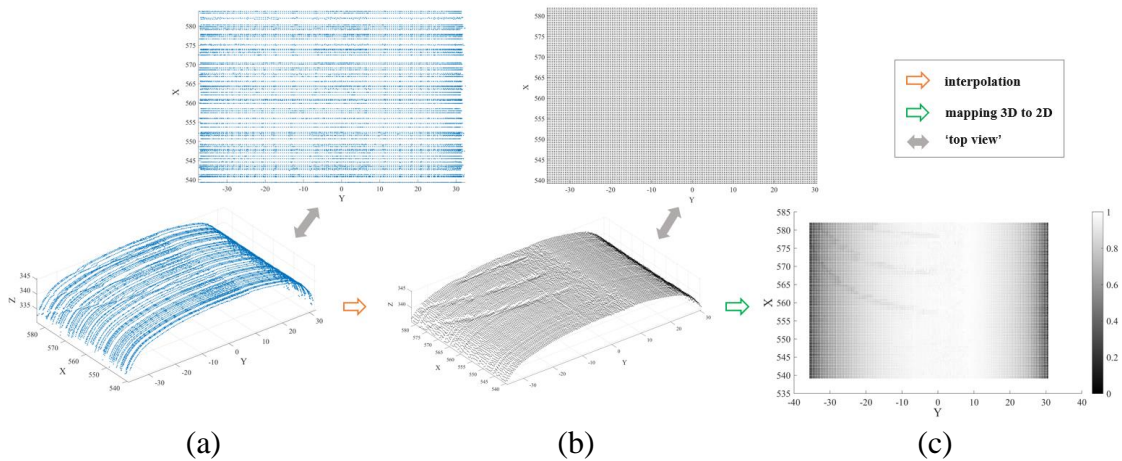
Fig. 6.5 Details of the mapping module for 3D inputs

information of each data point, while in the tensor representing laser sensor data, it relates to the depth (Z) of each data point.

As a result, the mapping module that consists of an interpolation layer and a mapping layer is used to transform 3D laser sensor data into 3D tensors with the same format as that of 2D images. Fig. 6.5 (b) and (c) visualise the transformation process of the sample. Because the proposed SMS involves manual movement of the laser sensor, 2D profiles that form the 3D data are normally distributed non-uniformly along the longitudinal axis of the rail due to non-uniform motion. In comparison, 2D images that fit the segmentation module are normally of a fixed size and the pixels that constitute the image are distributed uniformly in the 2D space, which eases the following training process. To restructure the 3D laser sensor data points, the interpolation layer based on the formula (6-3) is used.

$$Z = F(X, Y) \qquad (6\text{-}3)$$

The fundamental of the interpolation layer is the same as that of the interpolation method introduced in section 4.2.2-(2). The interpolation layer works by generating an interpolating surface (F) in space that encloses the raw data points first. The interpolating surface can then be used to interpolate a set of structured points with a specified density (e.g, $128(X) \times 128(Y)$ in this example). Finally, each interpolated point retrieves its depth by averaging the depth of raw data points in the nearest triangle. As shown in Fig. 6.5 (b), this produces an output with data points distributed uniformly in the space while also preserving the geometry of the original input.

As previously mentioned, the main difference between the tensor of the 2D image and 3D laser sensor data in this research is that their third dimensions relate to colour and depth information, respectively. Accordingly, a mapping layer is defined following the interpolation layer. The mapping layer works by converting the depth of each laser sensor data point into a depth-related colour intensity (Z_c). Since the distance between the SMS and the sample can vary during measurement, this results in the variation of the depth range of each sample. As the diverse range can increase the computational cost of subsequent procedures, a normalisation process defined by (6-4) was included in the mapping layer.

$$Z_c = \frac{Z - min\,(Z)}{ma\,x(Z) - min\,(Z)} \tag{6-4}$$

where Z represents the depth of laser sensor data points of a sample, and Z_c represents the normalised values. This ensures that all samples are homogenous with the depth-related intensity values (Z_c) ranging from 0 to 1. As visualised in Fig. 6.5(c), the output of the mapping module is a single-channel depth image that can be represented as (X, Y, Z_c). It preserves all of the information contained in the original input and also has the same properties as a high-quality image. More importantly, the colour information is derived from the depth of each point which is less affected by surface texture changes and contaminations, and thus can reflect surface conditions more accurately compared to conventional 2D images. The regions that correspond to the defects have a clear contrast against the background which makes defect detection easier.

(2) Segmentation module

Following the acquisition of high-quality depth images, this research aims to achieve pixel-level rail surface defect detection by accurately classifying each pixel in the corresponding image as healthy or part of a defect. This process is also known as semantic segmentation and thus the module here is called the segmentation module. As shown in Fig. 6.4, this research used an FCN in the U-Net architecture [104]. U-Net follows a standard encoder-decoder structure, which is explained in detail in the following two paragraphs.

The encoder can be divided into five blocks. Each block consists of a certain number of 3×3 convolutional layers with rectified linear units (ReLU) and a 2×2 max-pooling layer. The encoder with these successive layers works like a distillation filter. It extracts feature maps from the input images in a hierarchical manner. As shown in Fig. 6.4, the size of the feature map decreases while its depth increases along with the encoder. Therefore, the feature maps from the shallow layers (e.g., from blocks 1 & 2) are at higher resolution and typically contain more information about the location and boundaries of the object. The deep feature maps (e.g., from blocks 4 & 5) with increased depth usually have more abstract semantic information, which is useful for category classification.

The decoder is built up with 3×3 convolutional layers and up-sampling layers that are symmetric to the encoder. Feature maps go through the decoder and come out with decreased depth and increased size. This hierarchically reconstructs the spatial information and ultimately outputs probability tensors (segmentation maps) that denote the class of each pixel in the input images. It is worth noting that a batch normalisation layer is added between each 3×3 convolutional layer and up-sampling layer in this study. Batch normalisation is a technique that helps to standardise the input to a layer and thus reduce the computational cost of the training process. During the process of outputting segmentation maps, concatenation operators are used to hierarchically combine feature maps in different levels between the encoder and the decoder. This is normally considered to be the major advantage of the segmentation module using U-Net architecture, because such a combination enables the simultaneous use of global location and context, as well as preserving the full context information of the input images when generating the segmentation maps. Therefore, when compared to other models such as FCN8 [53] and SegNet [56], a model using U-Net architecture outperforms in terms of detail recovery. This will be demonstrated in section 6.2.5.

(3) Mapping module for 3D output

The outputs from the segmentation module are 2D segmentation maps, each of which is in essence a 2D matrix that contains the predicted class for each pixel (Fig. 6.6 (a)). In original 2D segmentation tasks, 2D segmentation maps are generally projected onto corresponding 2D images to denote the semantic information of the image. The proposed network in this study uses 3D laser measurement results as the input. This makes it possible to map 2D segmentation maps from the segmentation module back to corresponding 3D data, and thus achieve 3D segmentation. To achieve this, a mapping module that is symmetric to the input mapping module is added following the segmentation module. Fig. 6.6 visualises the process of mapping a 2D segmentation map (Fig. 6.6 (a)) back to its corresponding 3D measurement (Fig. 6.6 (b)). As a result, the network's output is 3D data points, with the points corresponding to the regions of defect highlighted. The 3D model of the rail sample (Fig. 6.6 (c)) and the defect itself (Fig. 6.6 (d)) can then be reconstructed by simply connecting or meshing the points together. This not only achieves pixel-level defect detection for the rail surface but can also help to achieve more comprehensive defect evaluation than conventional methods working in the 2D space alone.

(a)

(b)

Mapping 2D to 3D

'isometric view'

(c)

(d)
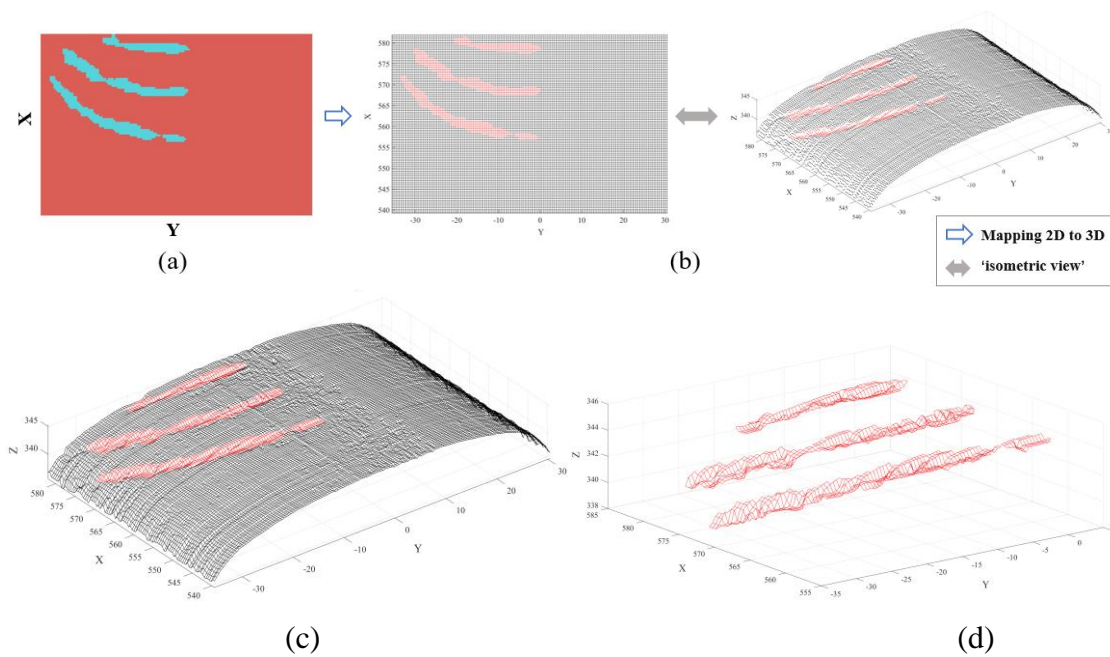
Fig. 6.6 Details of the mapping module for 3D outputs

6.2.3   Data preparation

With the dedicated network available, the next step is to prepare rail samples containing surface defects to build training and testing datasets, on which the network can be trained and tested. In this research, the preparation of data can be divided into two steps: (1) data collection and (2) data augmentation.

(1) Data collection

In deep learning, hundreds to thousands of samples are normally needed to ensure a normal learning process for a specific task. However, it is already known that rail surface defects are randomly distributed, and the in-service rail is currently mostly operated under healthy conditions with a rapid response to faults. Therefore, it is difficult to collect a sufficient number of defective samples in practice. Under this circumstance, this research selected a defect-free UIC 60 standard rail section (2000 mm (L) × 72 mm (W) × 172 mm (H)), on which hundreds of artificial surface defects were made in the workshop (Fig. 6.7(a)). These artificial defects represent three common types of surface defects (cracks, squats, and shelling) as has been introduced in section 2.1.1. Each artificial defect was made individually according to the representative of a fault type. Consequently, defects belonging to the same category have similar geometrical characteristics and, at the same time, each artificial defect has different dimensions from the others. This ensures the diversity and thus the

representativeness of these defects. Fig. 6.7 (b) shows some completed defective rail samples, with the quantitative information of the defects in each category summarised in Table 6.1.

<table>
<tr><td>(a)</td><td>(b)</td></tr>
</table>
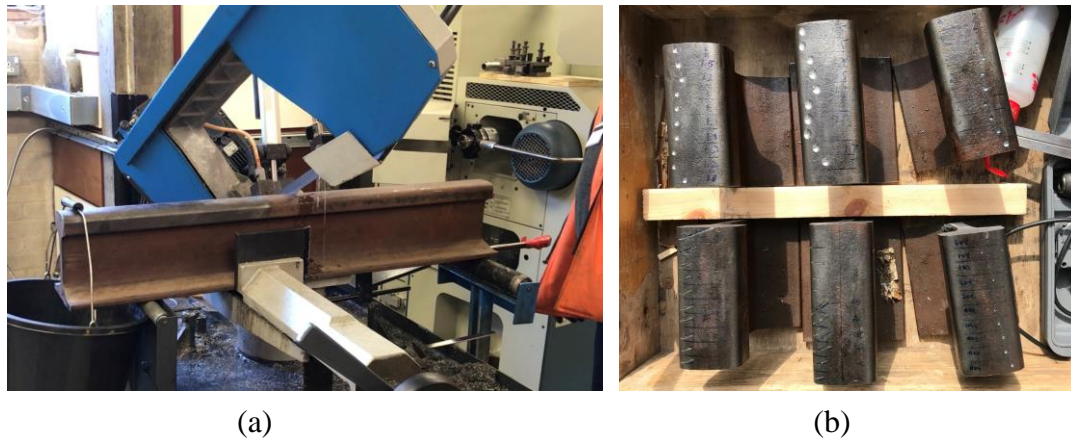
Fig. 6.7. Defective sample preparation: (a) machining a rail sample in the workshop; (b) some machined samples with artificial defects

Table 6.1 Quantitative information of rail samples and artificial defects

Defect	Surface length[1] (mm)	Maximum depth[2] (mm)	Surface angle[3] (°)/ width[4] (mm)	Cross-section view (approx. geometry)
Crack	8.0~27.0	1.20~5.60	90 (1.50 mm)	
Squat	6.10~9.70	1.10~1.70	6.0~9.20	
Shelling	1.95~5.30	2.80~5.30	1.80~5.30	

[1]Surface length refers to the size of the defect in the longitudinal direction of the rail.
[2]Maximum depth is the maximal distance from the bottom of the defect to the original surface of the rail.
[3]Surface angle is the cutting angle when making the artificial defects.
[4]Surface width refers to the size of the defect in the lateral direction of the rail.

These defective rails were then measured by the SMS in a random sectional manner as shown in Fig. 6.8 (a). Each measurement section generates a sample as shown in Fig. 6.8 (b). The random sectional measurement method further ensures the diversity of samples, some of which contain one defect or a partial defect, and others may contain multiple defects. Using this method, 100 defective samples were collected for each category. Considering the fact that a normal surface should also be expected during the inspection, another 100 normal samples were also collected from the normal rail sample. Therefore, 400 samples were collected, of which 300 are defective samples and another 100 are normal samples. All these samples were annotated at a pixel level to create the

ground-truth labels for training. These samples were then shuffled, and 75% of them (300 out of 400) were used as the training set and another 25% (100 out of 400) were used as testing set. Fig. 6.9 shows some of the samples along with their depth images and corresponding ground-truth labels for clarity.

(a)   (b)
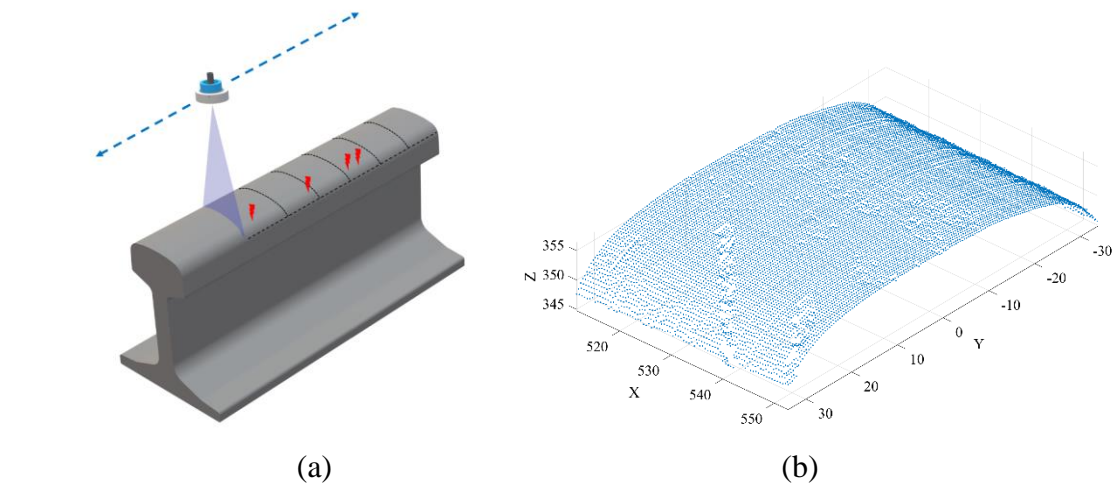
Fig. 6.8 Strategy of data collection: (a) schematic of the random sectional manner; (b) a measured sample

(a)

(b)

(c)

(d)
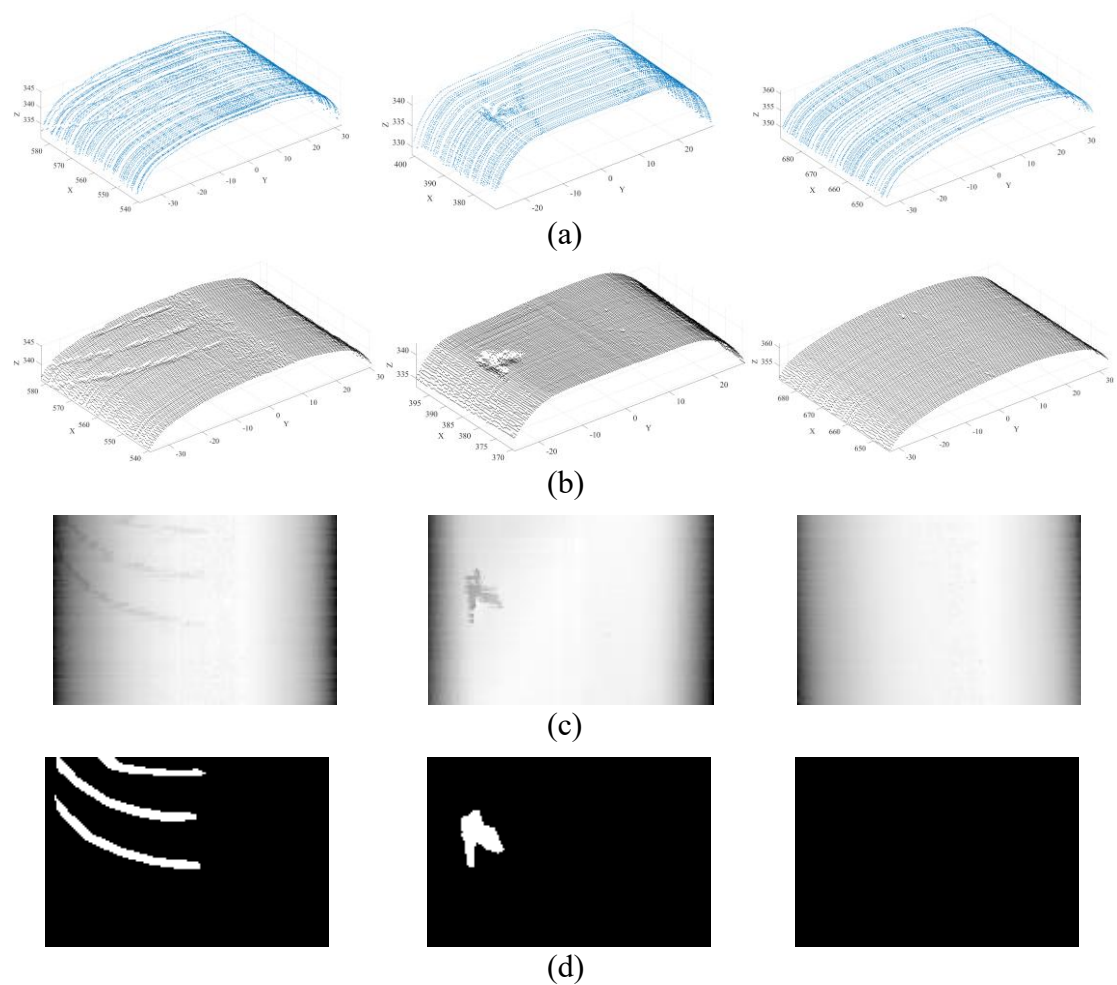
Fig. 6.9 Some examples of the training set: (a) raw dataset; (b) uniformly interpolated dataset; (c) depth-coloured images output from the mapping module; (d) ground-truth labels

(2) Data augmentation

In practice, the process of data collection in any form cannot avoid the effects of random noise due to factors such as environmental interference. However, data collection in the lab is normally less affected by noise, benefiting from the intentionally built working environment. This may yield a trained model that is not generic enough when facing new datasets captured from different environments. In deep learning for computer vision, data augmentation is a common method of generating more training data from existing training samples, by augmenting the samples via a number of transformations that yield slightly modified images. Data augmentation can increase the number and the diversity of training samples. Therefore, an appropriate pipeline of data augmentation can mitigate overfitting during the training process and help the trained model generalise better to new datasets by exposing it to more aspects of the data. For the proposed network, data augmentation can be applied to the depth images before they are fed into the segmentation module. In this research, different data augmentation methods were tested. A pipeline combining four different data augmentations was used in the final deployment, based on their different impacts on the training process of the model. Details of these augmentations are listed as follows:

- **Flip:** Horizontally flip input images.

- **Crop:** Randomly crop input images and resize the cropped images back to their input sizes.

- **Gaussian blur:** Add Gaussian blur with random intensity.

- **White Gaussian noise:** Add Gaussian noise to input images, sampled once per pixel from a normal distribution in the range of 0 to 1 (the range of the intensity value of the input images).

These augmentations were randomly combined and applied to the training set (both the depth images and corresponding ground-truth labels). Fig. 6.10 shows some examples of the augmented samples and the corresponding ground-truth labels. Theoretically, the combination of different augmentations can create an infinite number of non-repetitive training samples. However, since all the augmented samples came from a small number

of the original training set (i.e., 300 in this research), they are still heavily intercorrelated. Therefore, this research set the number of augmented samples to 600. Accordingly, the final number of samples for training is 900.

(a)

(b)

(c)
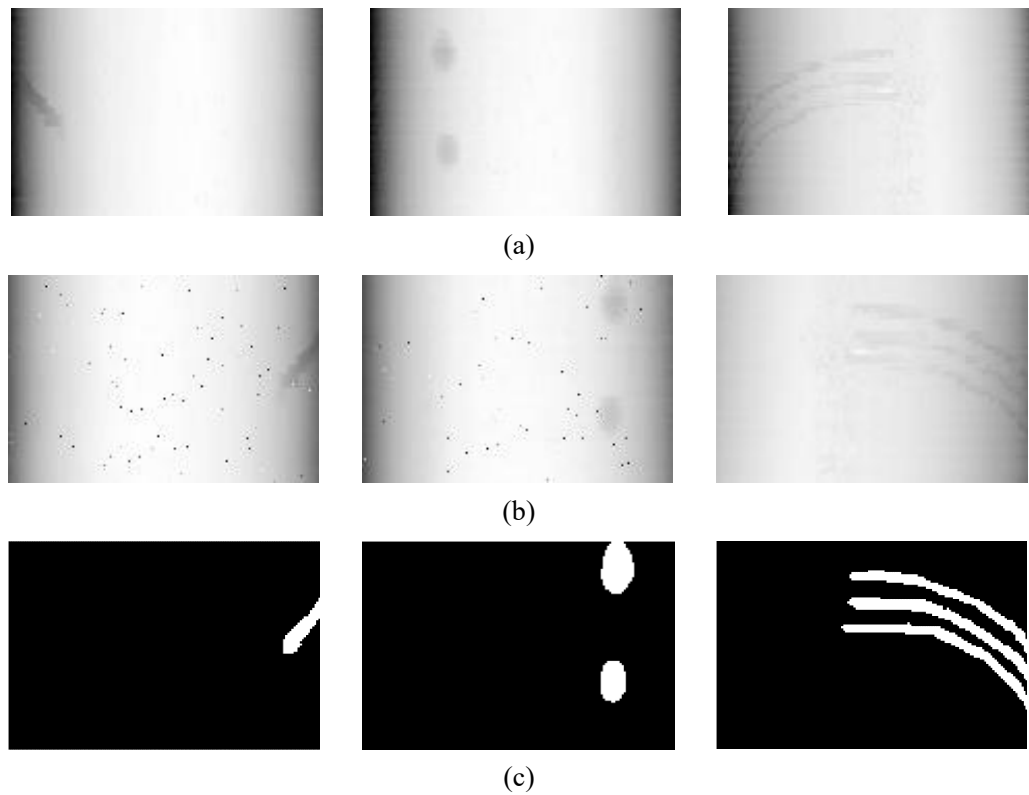
Fig. 6.10. Data augmentation: (a) original depth-coloured images; (b) augmented images; (c) synchronously augmented ground-truth labels

6.2.4 Strategy of network training and transfer learning

The previous section (6.2.2) introduced the architecture of the proposed network and the preparation of the dataset. To achieve pixel-level detection of rail surface defects, the network needs to be trained properly. This is essentially a process of optimising the weights assigned to each layer of the network to learn the transformation from the training samples (x_i) to their corresponding ground-truth segmentation maps (y_i). The flowchart in Fig. 6.11 illustrates the process of network training and its key procedures. Because the mapping modules in the proposed network have no trainable layer, the term 'forward propagation' refers to the process of outputting the predicted segmentation maps (y_i') through the encoder and the decoder. The loss score between each pixel of y_i and y_i' is then computed using the cross-entropy function defined by (6-5):

$$Loss = -\frac{1}{size} \sum_{i=1}^{size} y_i \cdot log\ y_i' + (1 - y_i) \cdot log\ (1 - y_i') \qquad (6\text{-}5)$$

where size is the number of scalar values (probabilities) in the output. Following the forward propagation, the term 'backward propagation' refers to the process of using the loss score as a feedback signal to update the weights of each layer in the encoder and the decoder. The weights are adjusted gradually using the stochastic gradient descent (SGD) method. Such forward and backward propagations iterate on a sufficient amount of data until the loss score between the predictions (y_i') and the true targets (y_i) meets the criteria. This yields a trained network or model that can perform well on new data sets, known as generalisation in deep learning.

To train a deep learning model from scratch (with random weights initialised), thousands of training samples are normally needed to help the model converge to meet the criteria. The dataset created here in section 6.2.3 is still relatively small even after the use of data augmentation. In this case, the model is prone to overfitting during the training process and can also result in a trained model with a low generalisation capacity. The concept of transfer learning is a common and effective method for deep learning on small image datasets. The basic idea of transfer learning is to train the model on a large and general dataset to generate a pre-trained model with transferrable features such as edge and boundary features. Because the pre-trained model has the initial
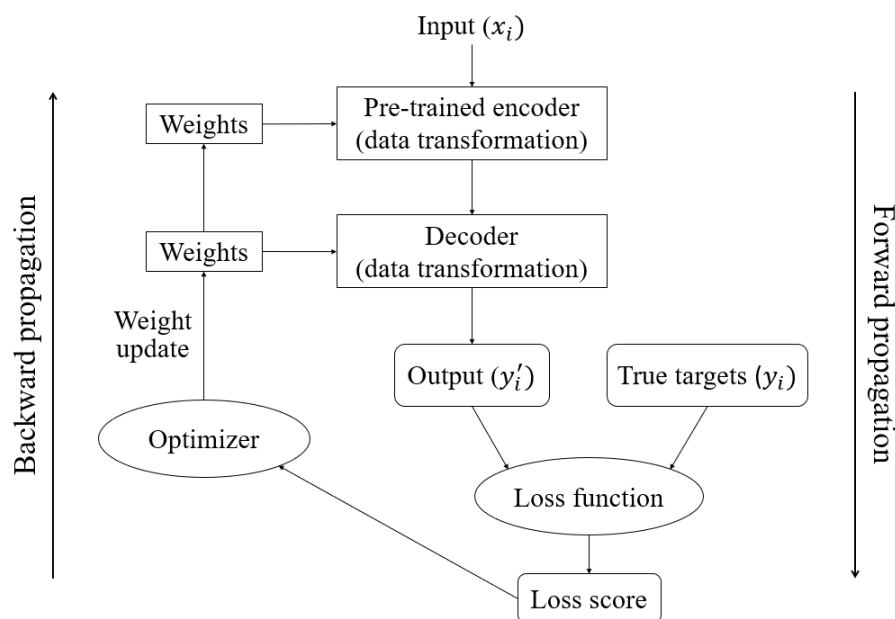
Fig. 6.11. Key processes of network training

weights learned from the large dataset, it can fit faster during the training process and also gain strong classification power and generalisation capability. Accordingly, as also indicated in Fig. 6.4, the concept of transfer learning is applied to this research by pre-training the encoder on the ImageNet data set that includes 1.5 million labelled images in 1000 different classes [105]. The pre-trained encoder can then be used and fine-tuned on the relatively small rail dataset during the training process, to fit the proposed segmentation task. Such a training strategy has been proven to be effective for improving both the accuracy and efficiency of the proposed network, which will be further discussed in section 6.2.5.

6.2.5 Experiments and results

The proposed defect detection strategy was evaluated on the dedicated 3D rail surface dataset. This section firstly introduces key processes of the experiment including setting parameters for the training process, the computation platform, and evaluation metrics. The training and testing results are then analysed. The optimised network used in this research adopts U-Net architecture as the segmentation module. To demonstrate the applicability of and the superiority of the network, comparisons with three other common segmentation networks (FCN8, SegNet, and PSPNet) were also considered with testing results presented and analysed.

(1) Parameter setting

As introduced in section 6.2.4, the initial weights of the encoder were pre-trained from the ImageNet dataset, and binary cross-entropy was used as the loss function. Apart from these, the optimiser used for updating and tuning the weights was SGD. The learning rate of the optimiser was set to 10^{-2} based on experiments, with a decay of 0.0016 and a momentum of 0.9 to optimise the learning process. When training, the dataset was divided into batches of 8. Each mini-batch contributes to a single gradient-descent update applied to the weights of the network. The network was iteratively trained about 5k times in 50 epochs.

(2) Computation platform

The network was built and evaluated in Python 3 using two main open-source platforms, Tensorflow and OpenCV. The hardware configuration was an NVIDIA Quadro P2200 GPU with 8 GB of memory on the Windows 10 operating system.

(3) Evaluation metrics

During the training process, the loss score calculated from the loss function (6-5) was used as an evaluation metric. Since pixel-level defect detection is in essence to classify each pixel to its category, classification accuracy was used as another metric. The loss score and accuracy quantify the overall performance of the model. To further evaluate the prediction results of the model, intersection over union (IoU) was used as the third metric [53]. IoU is commonly used for image segmentation tasks. It considers three pixel-level indices: true positive (TP), false positive (FP), and false negative (FN). Fig. 6.12 illustrates IoU using a person segmentation/identification task as an example. Among the subfigures, Fig. 6.12 (a) is the manually labelled ground truth (A). Fig. 6.12 (b) is assumed to be the prediction result (B) of the model. Then, the TP represents the number of pixels that are correctly predicted to belong to the person. It is equal to the intersection between A and B as shown in Fig. 6.12 (c). The FP represents the number
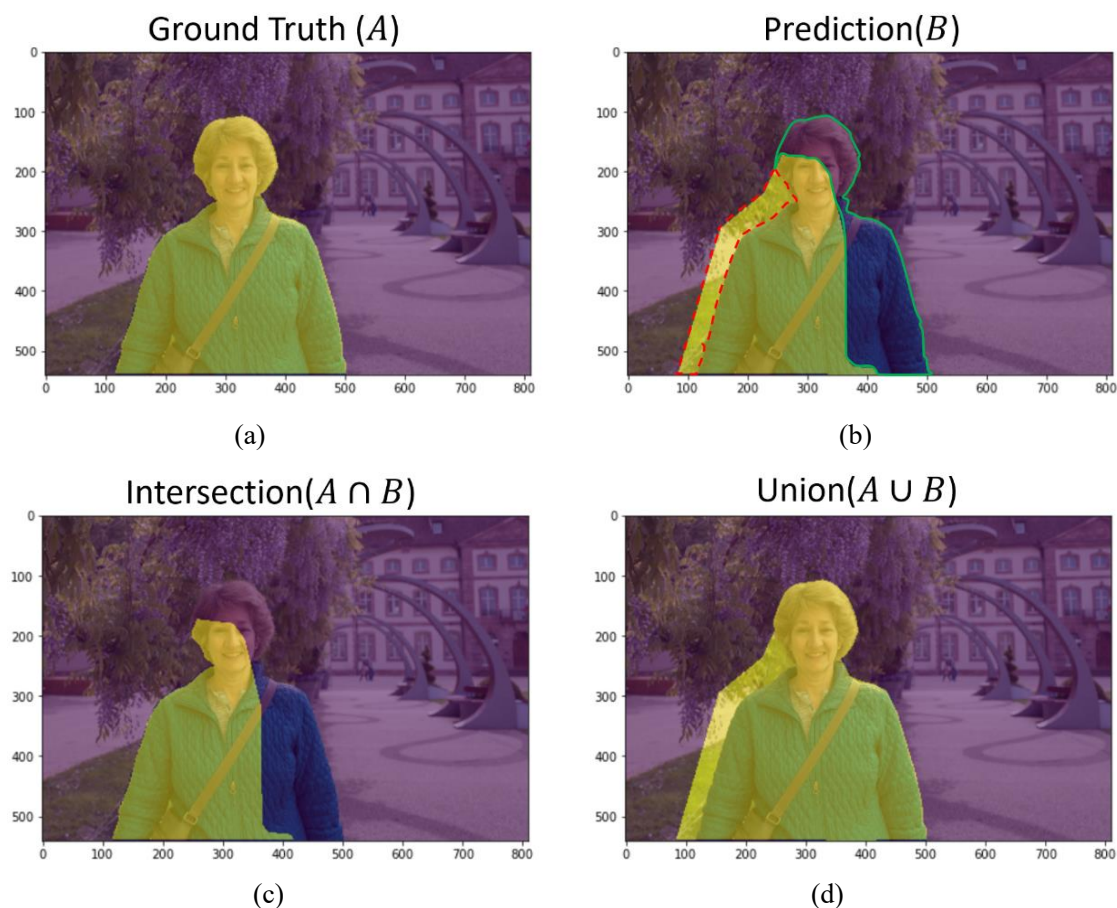
Fig. 6.12. Basics of IoU metric: (a) ground truth of a segmentation task; (b) prediction result; (c) intersection between ground truth and prediction; (d) union between ground truth and prediction

of pixels that are incorrectly predicted as a part of the person, which is highlighted with the dashed red line in Fig. 6.12 (b). The FN represents the number of pixels that are incorrectly predicted as part of the background, which is highlighted with the solid green line in Fig. 6.12 (b). These three indices (TP, FP, and FN) cover all possible situations of a pixel in a segmentation task of an image. Accordingly, the IoU is simply defined by (6-6):

$$
\begin{aligned}
IoU &= \frac{A \cap B}{A \cup B} \\
&= \frac{TP}{TP + FP + FN}
\end{aligned}
\tag{6-6}
$$

The IoU quantifies the degree of matching between the predicted segmentation maps and corresponding ground-truth labels. For a segmentation task, it should be expected that the larger the TP is, and the smaller the FP and FN are, the better the segmentation result.

(4) Experimental results and analysis

Tests of the proposed network and dataset: The proposed network contains a pre-trained encoder. To repurpose it for the inspection task in this research, there are primarily two methods: 1) retraining the entire network on the target dataset, or 2) retraining only the top layers of the encoder while training the other parts of the network. The second method is also known as fine-tuning, which is based on the shallow layers of a pre-trained encoder normally containing more generic and reusable features. Both of these two training strategies were used in this research. The fine-tuning strategy, by freezing the first three blocks of the encoder and jointly training blocks 4 and 5 of the encoder, outperforms other approaches in terms of the accuracy and speed of training and validation processes. The main reason is that the dataset in this research was relatively small and thus retraining the entire network had a higher risk of overfitting. Moreover, retraining the entire network involves many more parameters with increased computational complexity and thus needs a longer training process. Consequently, the fine-tuning strategy was applied in the final deployment. Fig. 6.13 visualises one training and validation process using the fine-tuning method. It includes the curves of training and validation accuracy, as well as the corresponding loss. The validation curves are closely tracking the training curves. The model's training and validation accuracy grow linearly with every epoch, while the losses

linearly decrease. The model's average validation accuracy remains consistently over 99.3% in all experiments.

(a)   (b)
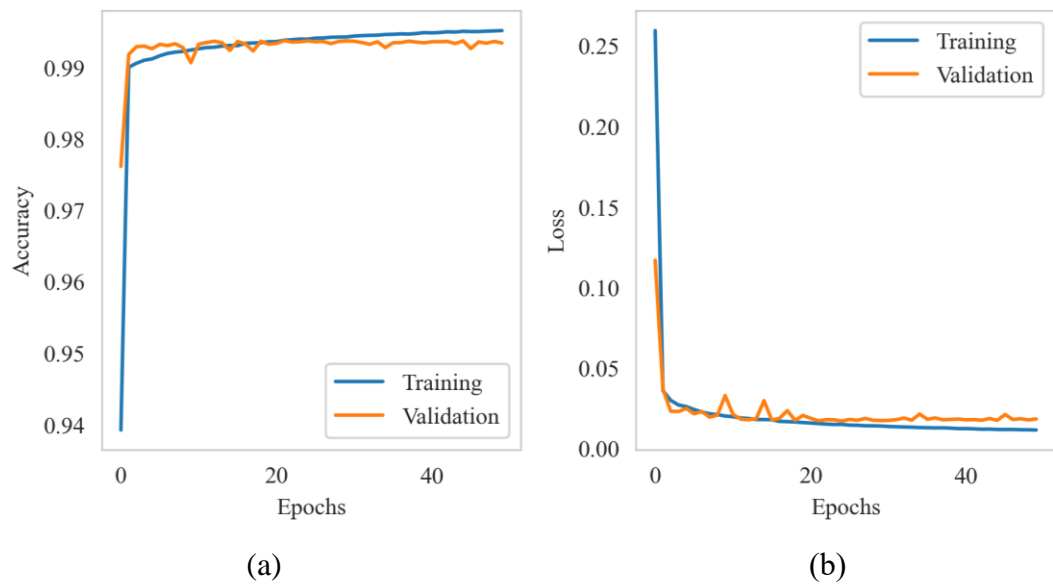
Fig. 6.13. Visualisation of a training and validation process

Fig. 6.14 visualises the output of the trained network using several testing samples as an example, among which (a) are the inputs and (b) and (c) are the corresponding depth-coloured images and ground-truth labels. Fig. 6.14 (d) shows the predicted 2D pixel-level segmentation maps from the network. The segmentation maps can then be mapped onto the corresponding input through the output mapping module (Fig. 6.14 (e)) to locate the defects and describe their boundaries precisely in 3D, on which the surface geometry of the rail can be reconstructed and characterised easily using a surface mesh (Fig. 6.14 (f)), to provide more comprehensive support for any further defect management and maintenance works, e.g., rail grinding.

Quantitative analysis of the pixel-level detection results for all testing samples is shown in Table 6.2. Because the majority of the surface defects in this research are at the millimetre level, the number of pixels that correspond to defects (TP+FN in the class of defect) only accounts for 2% of the total number of pixels (TP+FN in both classes). This proportion itself can pose challenges to pixel-level defect detection as most of the existing methods only consider defects with an area larger than 10% of the whole image. Under the circumstances, the proposed method can achieve a pixel-level defect detection with an accuracy of up to 78.5%, and its detection accuracy for other defect-free areas is as high as 99.4%, yielding a mean IoU (mIoU) of up to 87.9%. This result

is comparable to or even better than most existing pixel-level defect detection methods that rely on high-quality production line images. Meanwhile, the trained model needs only about 0.03 s on average to process each sample, which means it can achieve real-time inspection of more than 30 frames per second. Although the detection speed varies depending on the computational platform, it can still be used as a benchmark to evaluate the efficiency of the proposed method.

(a)

(b)

(c)

(d)

(e)

(f)
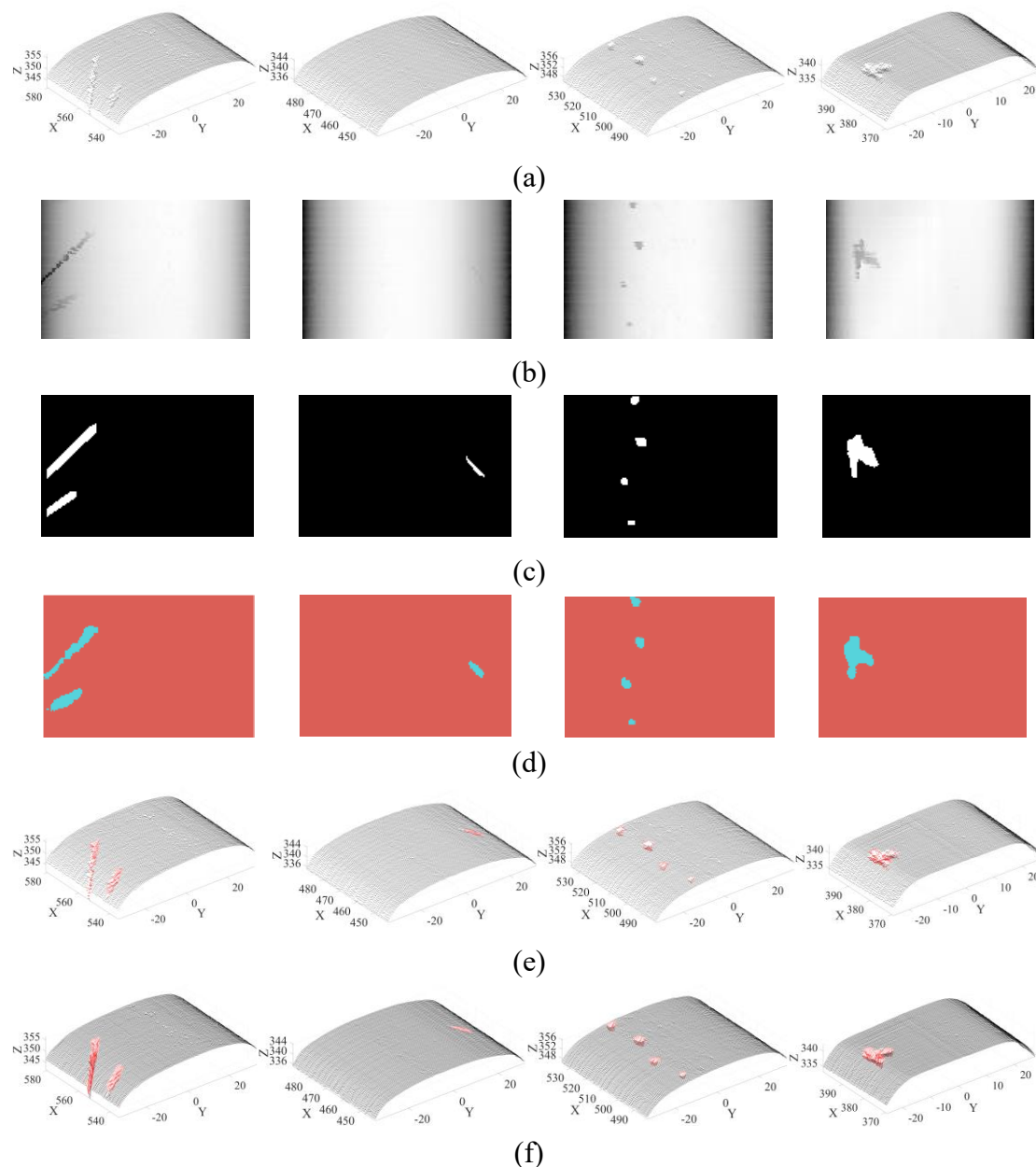
Fig. 6.14 Some detection results of a trained network: (a) input; (b) depth-coloured images; (c) corresponding ground-truth labels; (d) predicted 2D pixel-level segmentation maps; (e) output of the network; (f) reconstructed surface with defects detected

Table 6.2 Quantitative evaluation results

Class	TP (pixels)	FP (pixels)	FN (pixels)	IoU (%)
Defect	28749	3778	5062	76.5
Background	1420587	5062	3778	99.4
Mean IoU (%):	**87.9**	**Average speed (s): 0.03**		

Comparison with other segmentation models: The segmentation module in this research uses U-Net architecture. Benefiting from the modularised design of the structure of the network, any other 2D segmentation architectures can also be substituted as the segmentation module to adapt to different inspection tasks. Such flexibility along with the superiority of U-Net architecture was demonstrated by using three other popular architectures: FCN [53], SegNet [56], and PSPNet [106], as the segmentation module. It is worth noting that all these architectures use the same pre-trained encoder (VGG-16) but different decoding processes. Other implementation details such as the parameter setting and the computational platform are the same. Fig. 6.15 visualises some of the outputs of these modules on the same batch of testing samples, where rows (a) and (b) are the original input and corresponding ground-truth labels and rows (c) to (f) are the output of different segmentation modules. In this research, it was found that U-Net architecture outperforms the others in terms of both detection accuracy and integrity. Table 6.3 lists the quantitative comparisons of the pixel-level detection results for different architectures, which further demonstrates the superiority of the proposed network.

(a)

(b)

(c)

(d)

(e)

(f)
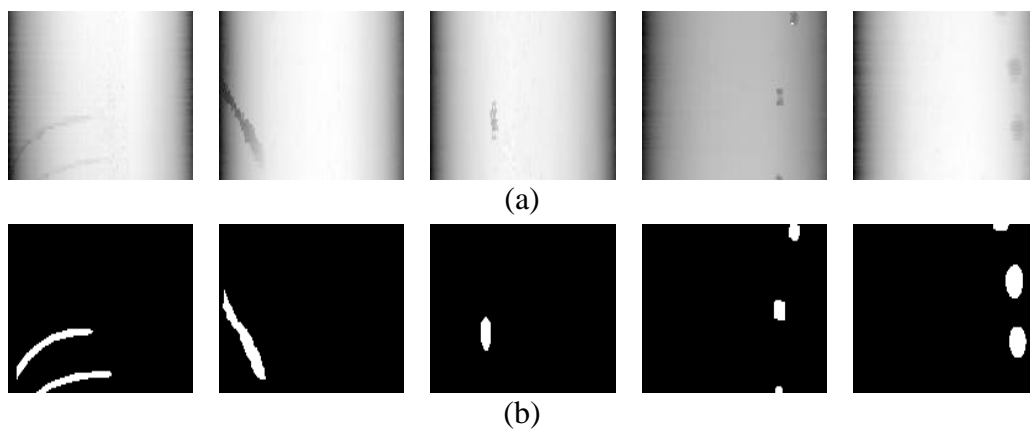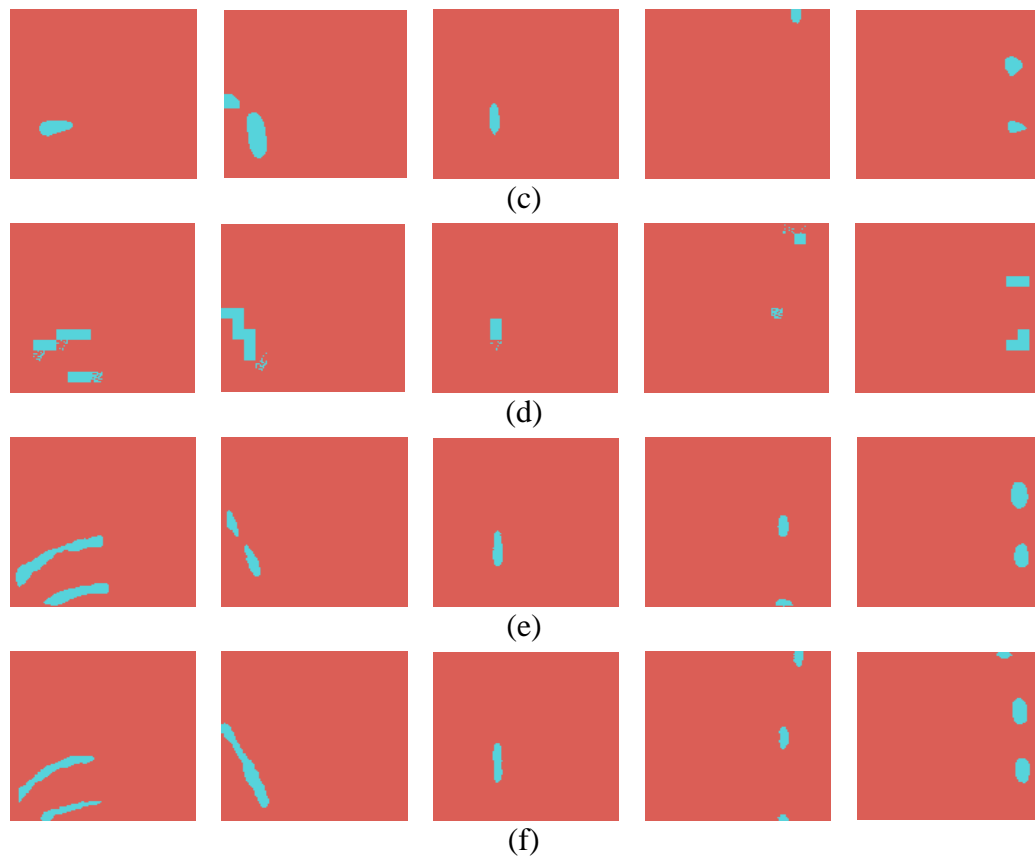
Fig. 6.15. Comparison results: (a) original input; (b) corresponding ground-truth labels; (c) PSPNet; (d) FCN; (e) SegNet; (f) U-Net

Table 6.3 Quantitative evaluation results

Segmentation module	PSPNet	FCN	SegNet	U-Net
mIoU (%)	75.3	74.1	83.6	87.9
Avg. speed (s)	0.03	0.04	0.04	0.03

6.3 Two-step 3D model-based defect classification and evaluation

The method introduced in the previous section, 6.2, achieves end-to-end 3D pixel-level defect detection. It can not only locate the defect and describe the defect boundary accurately but also allows for reconstruction of the rail and defect geometry in 3D, as shown in Fig. 6.16. Following defect detection, defect classification is often considered as a further step as they are two fundamental parts of a vision-based inspection task as has been discussed in subsection 2.2.2. Defect classification can further distinguish defects of different categories, which is helpful for further defect management and maintenance works.

It is already known that, in machine learning, classification is normally defined as a process to identify the mathematical regularity between the feature vectors (x) and classes (y) of the dataset. The feature vectors are a set of variables that contribute to the distributional difference of samples among each other in the feature space. The function $y' = f(x)$ that establishes the regularity is called the classifier [36]. To generate the feature vectors and build the classifier, there are two solutions: (1) is to add a classifier into the previous deep convolutional network to upgrade the binary classification problem that distinguishes the defects from the backgrounds to a multi-

(a)

(b)

(c)
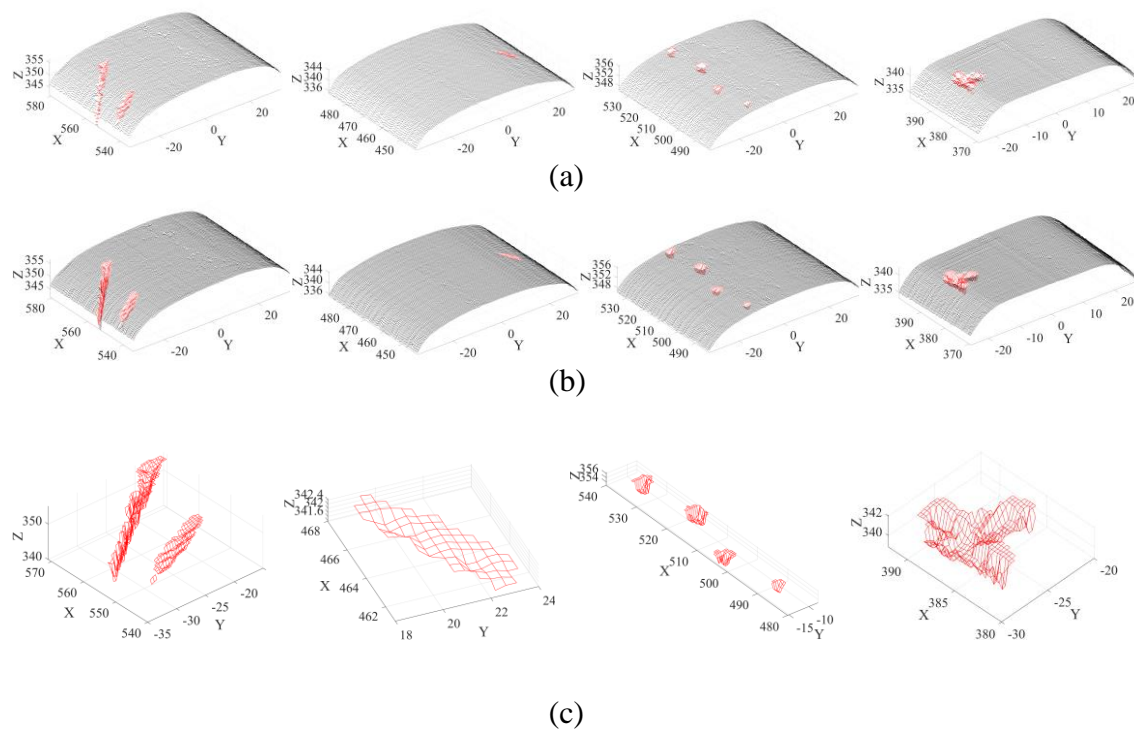
Fig. 6.16. Output of the proposed 3D pixel-level defect detection: (a) original output; (b) reconstructed output; (c) extracted defects

classification problem that also distinguishes between different instances of the defects, and let the network learn useful representations for classification during the training process; (2) is to build low-level feature vectors of the object based on human expertise, and then train a classifier separately. Since the dataset in this research is relatively small (400 samples in total), the first solution will inevitably increase the complexity of the original network and lead to the risk of overfitting, which can affect the network's defect detection performance. Meanwhile, the classification here only involves three different categories of surface defects, and the detected surface defects in 3D from the previous method already show a lot of geometric discrepancies among each category. These geometric discrepancies can be extracted easily as feature vectors. Therefore, the second solution was adopted in this research for the classification of surface defects.

6.3.1 Generic feature extraction

In conventional visual inspection, human inspectors normally measure the geometry of defects using contact measurement tools and carry out manual classification based on their geometric discrepancies such as shape and size, etc. The 3D pixel-level defect detection results from section 6.2 make it possible to quantify this information from the detected defects directly and use them as feature vectors to carry out the classification automatically. The features of length (L), width (W), and depth (D) that are related to the shape of the defects were considered first in this research. Fig. 6.17 visualises these features using a detected surface crack as an example. Fig. 6.17 (a) is the 3D view of the detected defect, and (b) is its top view. It can be seen that the defect is reconstructed with quadrilateral mesh elements having the same size, which is decided by the interpolation density mentioned in section 6.2.2-(1). This makes it simple to calculate the L and W of the defect by multiplying the number of mesh elements by the size of the mesh element in the corresponding direction. Similarly, the D of the defect is easily accessible as indicated by its side view in Fig. 6.17 (c). It is worth mentioning that in this research, the maximal L, W, and D in the corresponding directions (Y, X, Z) are measured to ensure feature measurement consistency across all samples.

(a)

(b)

(c)
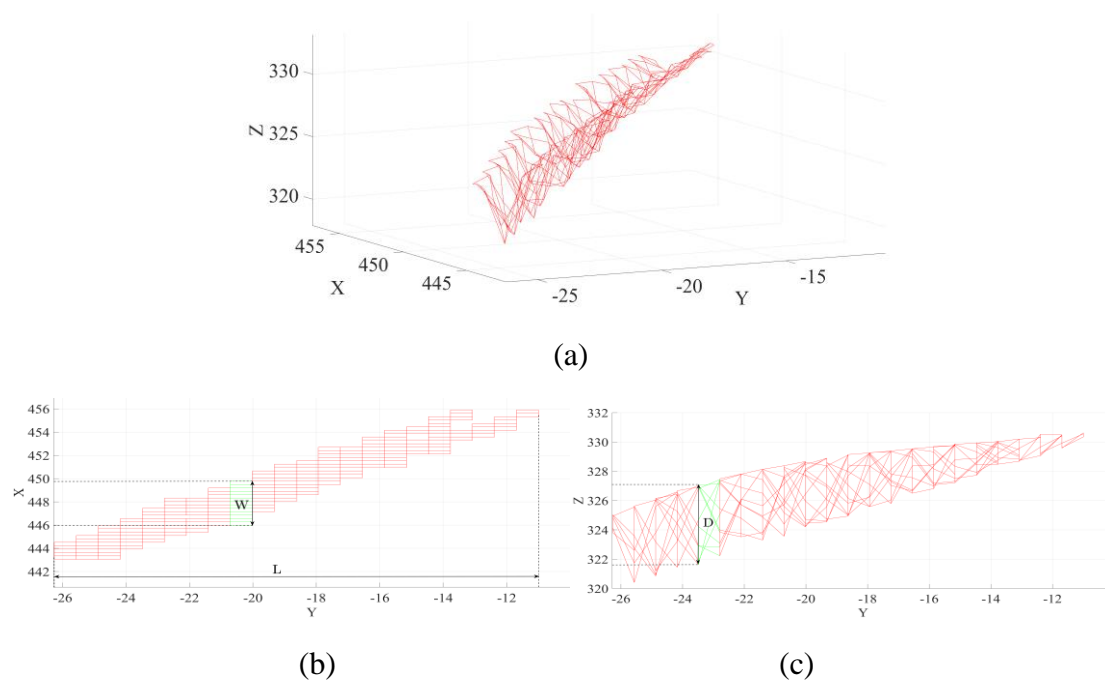
Fig. 6.17. Visualisation of the informative features

In addition to defect classification, the evaluation of intra-class differences like the severity is also critical for directing appropriate maintenance operations such as rail grinding or replacement. In practice, such works normally rely on human observers' knowledge to assess the size of the detected defect, which can be time-consuming and subjective [79]. The detection and characterisation of defects in 3D allow for not only the previously mentioned 2D size information (L, W, and D) but also more comprehensive 3D size information. For example, volume is one such metric that can reflect the severity of defects and their impact on the rail directly. This information is not accessible in conventional inspection methods with the output in 2D. Therefore, the volume of detected defects was also considered as a further feature vector. The challenge posed here is that all the detected defects have irregular spatial structures. One common solution for this is to divide the irregular geometry into a set of smaller parts that have regular geometry [107]. The volumes of these smaller parts are easy to compute or approximate. Therefore, the volume of the entire object can be approximated by summing up the volumes of these simpler dissections. Fig. 6.18 illustrates the principles of this method using a detected rail surface crack and a detected rail surface squat as examples. Fig. 6.18 (a) and (c) are the top view of 3D models of the defects about the X–Y plane. The quadrilateral mesh elements allow the volume of

defects to be divided into a set of rectangular solids. The depth of these rectangular solids can be approximated by the rectangular boxes in the side view of 3D models of the defects, as indicated in Fig. 6.18 (b) and (d). The linear nature of the rail determines that all the mesh elements in the same column about the Y-axis have the same reference height to compare, which can be derived from the edge mesh elements of the 3D model. As a result, the volume of each defect with irregular geometry can be approximated by these rectangular solids with known bottom areas and depths. If we use V_{solids} to represent the total volume of rectangular solids that form a defect, the volume (V) of this defect can be approximated when $\Delta l \rightarrow 0$ and $\Delta w \rightarrow 0$ with an increase of k on the basis of the definite integral principle.

$$V_{solids} = \lim_{\substack{\Delta l \to 0 \\ \Delta w \to 0}} \sum_{k} \Delta l \times \Delta w \times D_k$$

$$= V$$

(6-7)

where Δl and Δw are the length and width of the mesh element, and D_k is the depth of the corresponding mesh element.

(a)

(b)

(c)

(d)
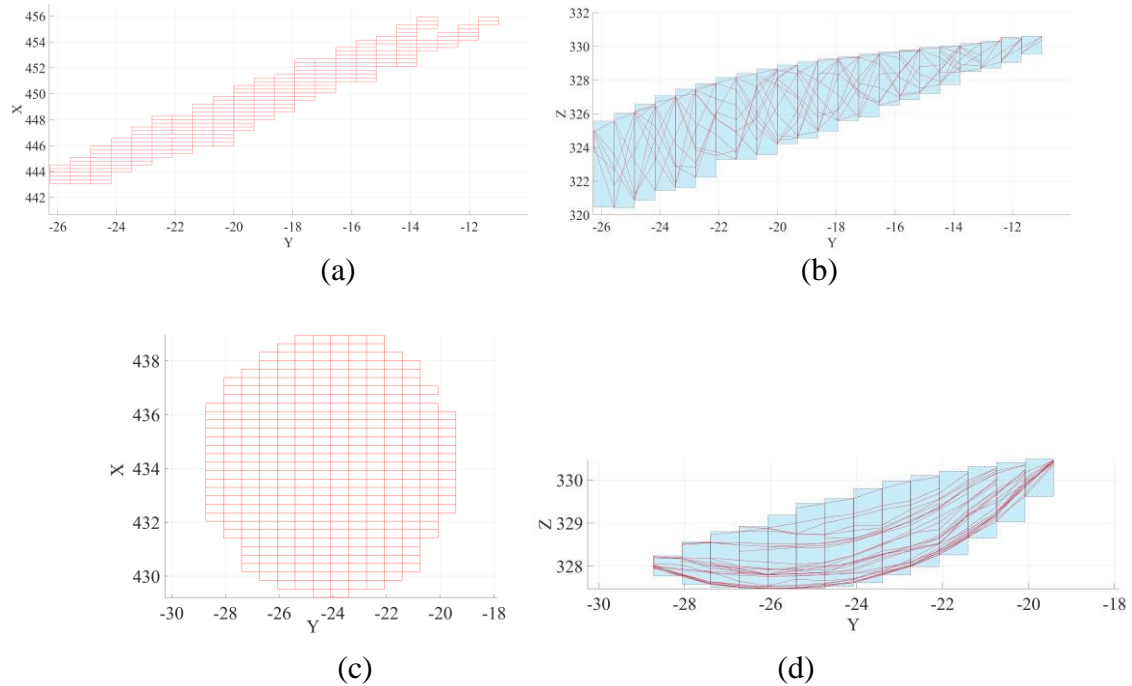
Fig. 6.18. Volume approximation of the defect

In summary, four quantified geometrical features (L, W, D, and V) can be generated from each 3D surface defect, which build distinguishable feature patterns to classify detected defects into the corresponding category. This strategy was evaluated by

randomly selecting 100 detected surface defects from the previous section 6.2. All 100 detected defects generate a 100×4 feature set. For clarity, a subset of the full feature set including an example from each defect type and severity is shown in Table 6.4. In addition to the three categories of rail surface defects that were considered in this research (crack, squat, and shelling), each category was further divided into two levels of severity (L1 (minor) and L2 (more significant)) based on expert knowledge. Further processes on the training and evaluation of the classifier are introduced in the following sections 6.3.2 and 6.3.3.

Table 6.4 Subset of the full feature set for 3D model-based rail surface defect classification

Feature / Class	Length (mm)	Width (mm)	Depth (mm)	Volume (mm^3)
1. Crack_L1	15.34	1.70	1.80	23.47
2. Crack_L2	24.96	2.70	4.30	144.89
3. Squat_L1	6.36	6.20	1.0	8.82
4. Squat_L2	9.91	9.30	1.60	34.33
5. Shelling_L1	3.54	3.30	1.90	17.43
6. Shelling_L2	5.62	5.10	4.60	103.55

6.3.2 Design of the classification strategy

With the feature set available, the next step is to build the classifier (f) to establish the mathematical regularity between the feature vectors (x) and classes (y) of the dataset.

$$y' = f(x)$$

This thereby can potentially reduce the dependency on human experts, to carry out surface defect classification and evaluation automatically.

Nowadays, a large number of classifiers have been developed, from Bayesian to the latest neural networks. Therefore, one essential activity is to select appropriate classifiers. For a relatively low dimensional feature set (e.g., 4 in this research), the most common and efficient method is to visualise the distribution characteristics of the features from all classes in the spatial domain, for example, to see whether they are linearly separable or not. This is a key index for most classification models, and thus can be used to assist the selection of appropriate classifiers. For visual clarity, two

random features from all six classes are visualised using the scatter plot method, as shown in Fig. 6.19. Data points with the same shape represent defects in the same class. It can be observed that these defects are linearly separable in some dimensions, with relatively clear margins between each other as shown in Fig. 6.19 (a), but are linearly non-separable in other dimensions as shown in Fig. 6.19 (b). This characteristic of distribution makes SVM-based classifiers good candidates given their relatively low computational complexity and flexible definition of hyperplanes. For linearly separable datasets, the hyperplane in SVMs divides them into different classes while allowing the maximum margins. Furthermore, the introduction of kernel functions in SVMs allows linearly non-separable datasets to be converted to be linearly separable ones by mapping them into a higher dimensional space. Therefore, SVMs were selected as the candidate classifiers in this research.

(a)   (b)
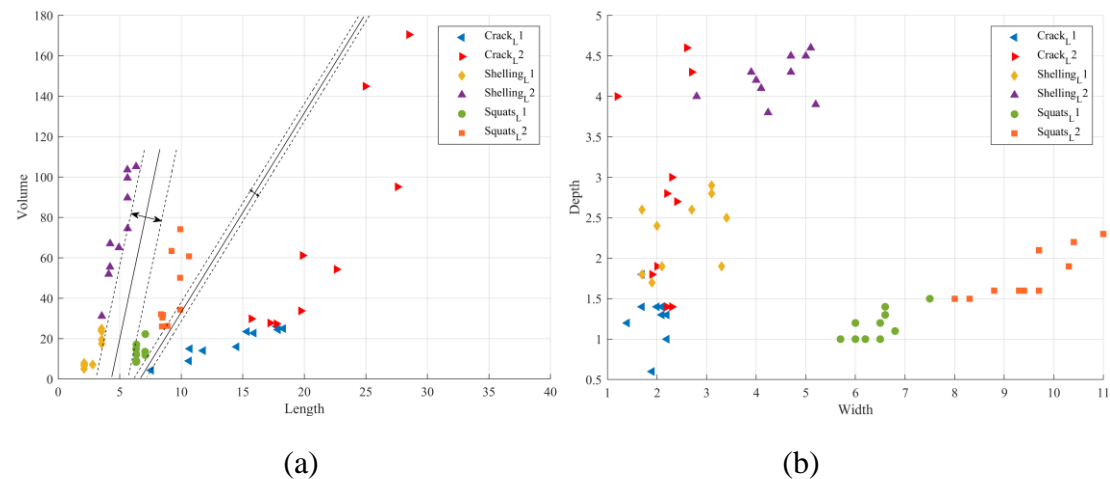
Fig. 6.19. Space distribution characteristics of the full feature set: (a) distribution by length and volume; (c) distribution by width and depth

6.3.3   Experiments and results

The proposed 3D model-based surface defect classification method was evaluated on a dataset containing 100 surface defects that are randomly selected from the previous section 6.2. This section introduces the key processes of the experiment and analyses the experimental results. The computation platform was the same as the one introduced in 6.2.5-(2), so it is unnecessary to go into detail here. The other two key processes of the experiment, setting the multi-class SVM parameters and evaluation metrics, are introduced firstly. The experimental results are then analysed in detail. As mentioned earlier in section 6.3.2, there are many other classifiers. Comparison with the other two common classifiers (KNNs and FNNs) was also considered, with comparison results presented for reference.

(1) Parameter setting

For SVMs, the key parameters include the regularisation parameters (C) and the kernel coefficient (γ) [108]. In this research, multi-class SVMs including a linear SVM and non-linear SVMs with quadratic, cubic, and Gaussian kernels were considered. The regularisation parameter (C) of all these SVMs is set to 1. The kernel coefficient (γ) is set to *1/F* for the quadratic kernel and the cubic kernel, where F is the number of predictors (features) and is equal to 4 herein. As the Gaussian kernel is most commonly used, it was typically tested on three different kernel scales: $\sigma = \frac{\sqrt{F}}{4}$, $\sigma = \sqrt{F}$, and $\sigma = 4 * \sqrt{F}$. Here, σ has the following relation with γ:

$$\gamma = \frac{1}{2 \cdot \sigma^2} \tag{6-8}$$

according to the Gaussian kernel function defined by (6-9):

$$k(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

$$= exp\left(-\gamma \cdot \|x_i - x_j\|^2\right) \tag{6-9}$$

Additionally, due to the different scales of geometric features, data normalisation is considered in the form of $x_j = x_j - \mu_j$, where x_j is the original value of each feature and μ_j is the mean value of the specific type of feature, to improve the fit. The basic

logic of the proposed 3D model-based rail surface defect classification method is summarised in Fig. 6.20.
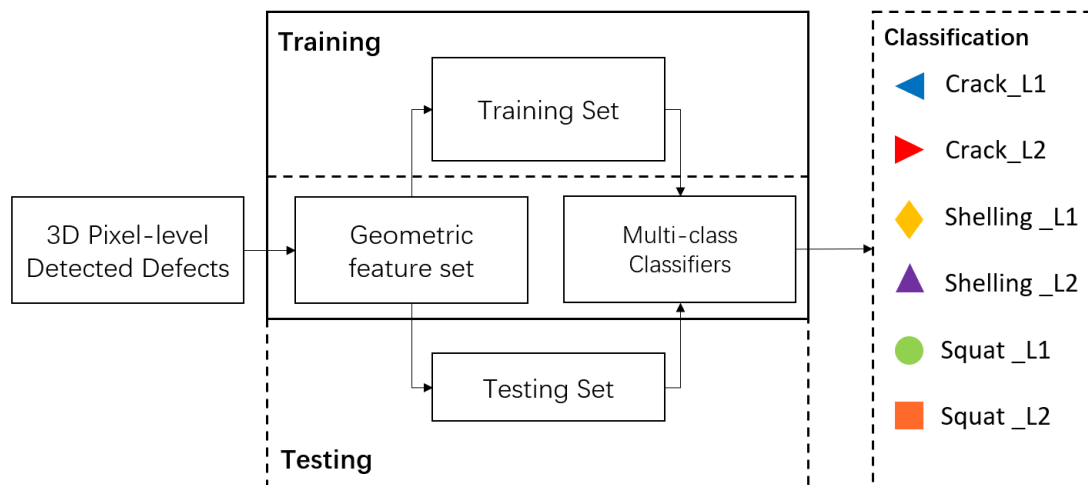
Fig. 6.20. 3D model-based defect classification block processing

(2) Evaluation metrics

This research uses three evaluation metrics to access the performance of the proposed surface defect classification approach: (1) the classification accuracy, (2) the variance of k-fold cross-validation, and (3) the prediction speed. The classification accuracy was used to quantify the proportion of testing samples that are correctly classified. Since the classification task involves multiple classes (not a binary classification) and the dataset is small (i.e., 100 in this research), a simple training and testing split may give a smaller testing set that is not representative. For example, the samples in the testing set may not cover all six classes, resulting in a biased testing result. Accordingly, we used a strategy known as cross-validation in this research to solve this problem. In a cross-validation process, the full dataset is randomly split into k subsets. Each subset will be retained as a testing dataset, while the remaining $k-1$ subsets will be used as a training dataset. This process will be iterated throughout the whole k folds. In each fold, the testing dataset will be hidden from the model during the training phase, which means that each sample can be used for testing once. The k-fold cross-validation will derive k evaluation scores (E_i), which can then be averaged to assess the classification accuracy of the classifier, giving a less biased result [109]. The variance of k scores was used to evaluate the stability of the classifier. Additionally, the prediction speed was analysed in the form of observations per second. Although the prediction speed may vary depending on the computational platform, it can still be used as an index for the efficiency of

different classifiers. The diagram in Fig. 6.21 illustrates the process of *k*-fold cross-validation, where k is set to 5 here to ensure that each subset is large enough to be statistically representative of the full dataset.
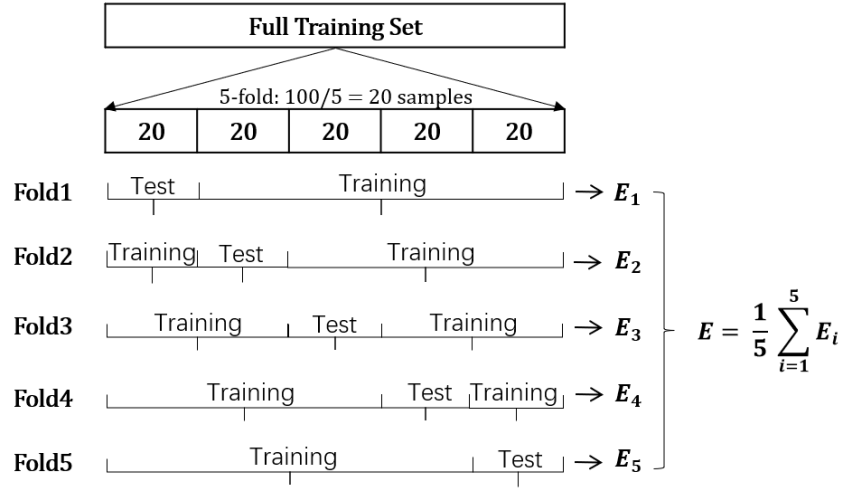
Fig. 6.21. k-Fold cross validation

(3) Experimental results and analysis

Test of the feature set and classifiers: All the classifiers were assembled and tested using the classification toolbox in MATLAB, which provides a fast and convenient environment for testing different classifiers. Table 6.5 summarises the testing results of all SVMs and the best results from KNNs and FNNs. For the KNNs, the best result is for the one when the number of closet neighbours is set to 10. For the FNNs, a three-layer neural network with a $6 \times 10 \times 6$ structure (i.e., 6 input features $\times$ 10 hidden neurons $\times$ 6 outputs) gives the best result. In total, the classifier with a Gaussian kernel ($\sigma = \sqrt{F}$) gives a classification accuracy up to 96.7% with no variance and a prediction speed of around 1200 observations/s, which outperforms other classifiers in this case. In comparison, FNN gives an average accuracy of 95.0%, while it suffers from a relatively high bias due to the overfitting on a small dataset. This could be improved with increased data volumes in future work. Accordingly, following the 3D pixel-level defect detection introduced in section 6.2, the detected defects can be further classified and evaluated based on their geometric features. Fig. 6.22 shows some examples of the output. Fig. 6.22 (a) are the original defect detection results. Fig. 6.22 (b) are the classification results. Fig. 6.22 (c) are the extracted defects with different

colours indicating their corresponding classes. All of these provide essential references for further defect management and maintenance works.

Table 6.5 Comparison of the performance of multi-class classifiers on the 3D-model based feature set

Kernel	Full feature set		
	Cross-validation accuracy (%)	Score variance (%)	Prediction speed (obs./s)
Linear SVM	91.7	1.34	~590
Quadratic SVM	91.7	0.00	~910
Cubic SVM	88.3	0.66	~1200
Gaussian SVM ($\sigma = \sqrt{F}/4$)	71.7	1.34	~1100
Gaussian SVM ($\sigma = \sqrt{F}$)	96.7	0.00	~1200
Gaussian SVM ($\sigma = 4 * \sqrt{F}$)	85.0	0.68	~1200
Euclidean KNN ($k = 10$)	93.3	0.68	~900
FNN (6-10-6)	95.0	2.02	~850
Result	**96.7**	**0.00**	**~1200**

(a)

(b)

(c)
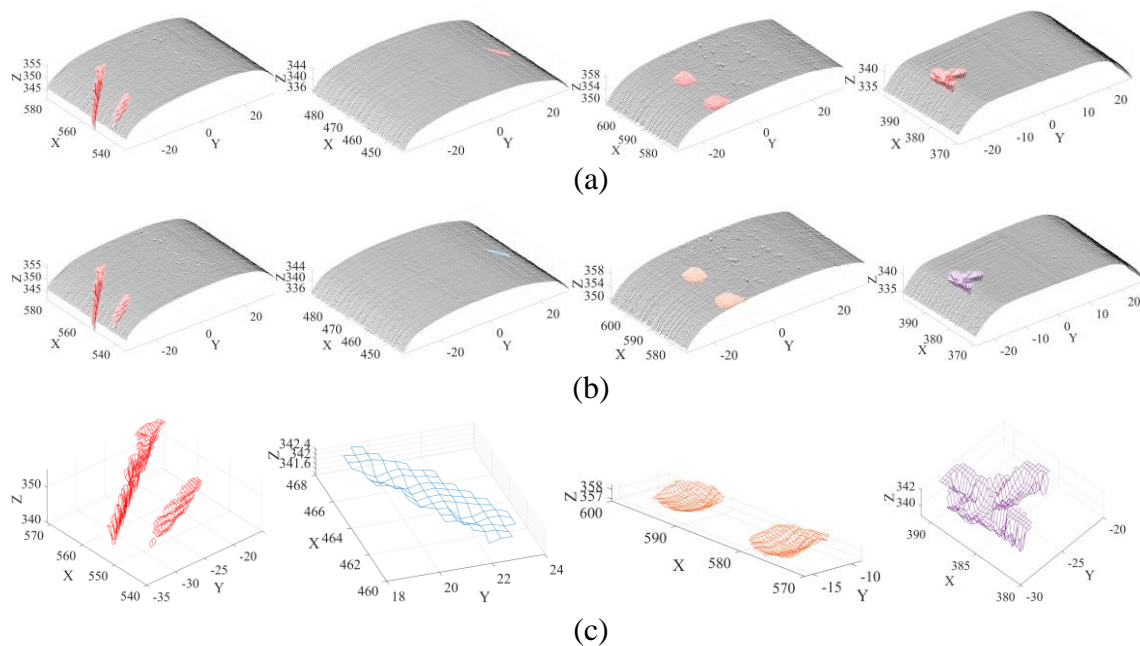
Fig. 6.22. Output of the proposed 3D model-based defect classification and evaluation: (a) pixel-level detection results; (b) classification results; (c) extracted defects after classification

Error analysis: To further analyse the performance of the 3D model-based defect classification method and find incorrectly classified samples that cause errors, the confusion matrix of the best classification result shown in Table 6.5 (i.e., Gaussian SVM ($\sigma = \sqrt{F}$)) is plotted in Fig. 6.23. The confusion matrix indicates the exact position of incorrectly classified samples and the positive classification rate for each class. It can be observed that only 20% (2 out of 10) of the samples in the true class of Crack_L2 are predicted as Crack_L1. All other classes are correctly classified with a 100% positive rate. Meanwhile, it is noteworthy that there is no miscategorised sample; all the incorrectly classified samples are just predicted to be at the wrong severity level. Further investigation of the incorrectly classified samples shows that the corresponding cracks are mainly distributed at the edge of the rail. The length of these samples can be measured incorrectly due to the shadowing effect of the laser line, as has been discussed in section 4.3, but the width and depth have relatively small measurement errors [110]. This can result in a sample measured with a length in the range of Crack_L1, but with other features such as volume in the range of Crack_L2, which could potentially mislead the classifier. This can be solved in future work either by improving the measuring unit or eliminating the influence of measurement errors in post-processing such as by adding confidence indices to the samples based on their geometry.
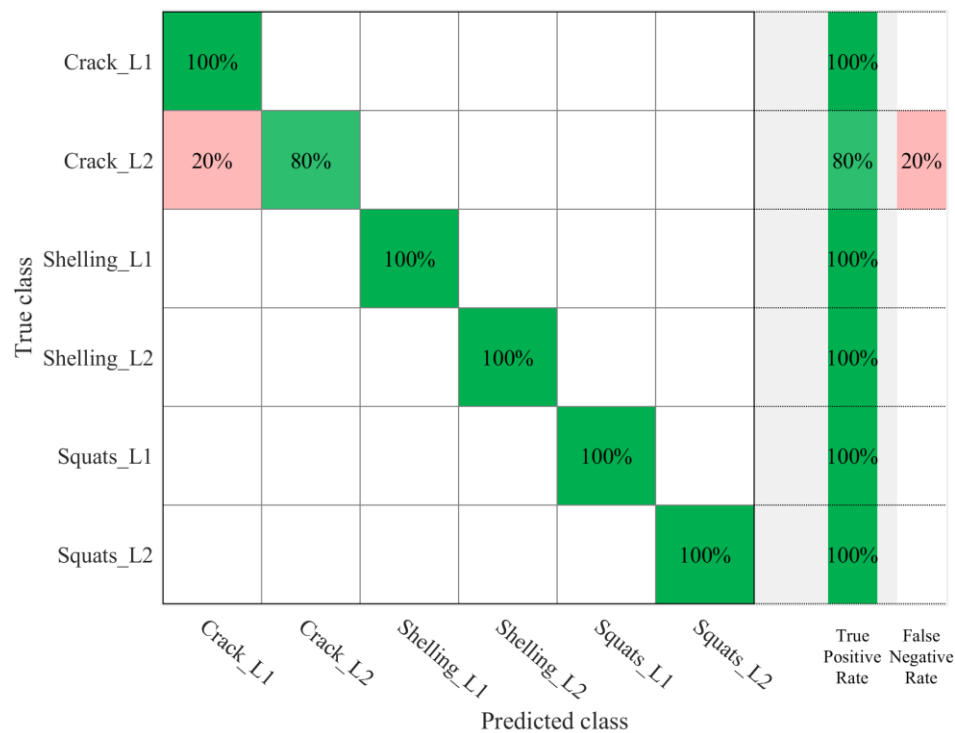
Fig. 6.23. Confusion matrices for Gaussian SVM classification result

6.4 Summary

High-precision laser triangulation sensors have their strengths in measuring surface geometries. Conventionally, in railway systems, they have been mainly used as 2D rail profile measuring units, which allows for limited inspection strategies. More refined inspection tasks such as defect detection and classification still rely heavily on human expertise. As has been introduced in Chapters 4 and 5, the proposed multi-sensing system in combination with dedicated data integration algorithms allows for not only better 2D rail profile measurement but also accurate and comprehensive characterisation of rail surface conditions in 3D. With 3D measurement results, this chapter introduces two new strategies for automatic rail surface defect detection and classification. For defect detection, transferability between 3D surface measurement data and 2D images is analysed; advanced deep learning models in computer vision are migrated to process high-precision 3D rail surface measurement data from the proposed multi-sensing system. An optimised deep semantic segmentation network is developed successfully, achieving pixel-level rail surface defect detection. Compared to existing surface defect detection methods using 2D image data, the proposed method can not only locate the defect and describe its boundary accurately, but also allows for reconstruction of the rail and defect geometry in 3D. For defect classification, four geometric features are extracted from the defects to provide a low dimensional description for each of them. Multiple classifiers are tested to further classify detected defects into three common classes and two severity levels.

The experimental results demonstrate the strengths of multi-sensing data from the proposed system. It not only provides a more accurate and dimensional measurement for the rail but also allows for more intelligent and automatic inspections. The optimised deep semantic segmentation network in combination with the concept of transfer learning and data augmentation can obtain end-to-end pixel-level defect detection. The average detection accuracy of the trained model is above 99.4 %, the pixel accuracy measured by mIoU is up to 87.9%, and the detection speed can reach $30 - 35$ fps/s on the computational platform used in this research. Similarly, the classification accuracy of the trained classifiers can be up to 96.7%, with a prediction speed up to 1200 observations/s. These trained models can potentially replace human inspectors or at least reduce the dependency on them when carrying out rail surface defect detection and classification works. The two inspection strategies introduced in this chapter further

broaden the application of the proposed multi-sensing system, and also provide solutions for the development of multi-sensing inspection strategies in other relevant areas.

7 CONCLUSIONS AND FUTURE WORK

In railway systems, rail inspection is of vital importance; it ensures the rail in service satisfies relevant quality standards, supports appropriate maintenance works, and eventually ensures the normal and safe operation of the whole network. The increases of train speed, axle load, and traffic density in modern railway systems have posed challenges to conventional rail inspection methods that are normally dominated by human inspectors because they are normally characterised by low efficiency and can be subjective. Benefiting from the development of advanced sensory technologies, new rail inspection approaches are developing towards digitisation, automation, and intelligence; they have attracted a lot of attention of both researchers and railway operators, with many new systems developed aiming to carry out rail inspection more accurately, efficiently, and safely with reduced human involvement and inspection cost. This thesis summarises the author's research on existing rail inspection methods, detailed investigations on the laser- and vision-based rail inspection techniques, and contributions of developing new and systematic multi-sensing inspection approaches by integrating laser, vision, and other NDT sensors. The research works presented in this thesis support the author's hypothesis that the performance of laser- and vision-based rail inspection can be improved by effectively combining the strengths of disparate sensors and eliminating their shortcomings. Specifically, the work described in this thesis can be divided into three stages:

(1) the identification and categorisation of inspection targets and the selection of laser and vision as candidate techniques (Chapter 2).

(2) the development of a vision-based switch rail condition monitoring and inspection approach (Chapter 3) and a laser-based rail inspection prototype (Chapter 4), to identify the principles, characteristics, and limitations of vision and laser technologies as two individual inspection techniques. This allowed the development of a new prototype for a laser-based multi-sensing rail inspection system (Chapter 5).

(3) the combination of vision and laser inspection algorithms on multi-sensing data to develop more intelligent surface defect inspection strategies (Chapter 6).

7.1 Stage one – Inspection targets and technologies

In carrying out rail inspection, the inspection scope and targets determine the inspection approaches that can be applied. Accordingly, stage one of this research reviewed different rail inspection and rail defect management standards, as discussed in section 2.1.1. This helped to categorise rail surface defects into three groups in this research: 1) cracks, 2) squats, and 3) shelling, based on their initiation, characteristics, and distribution. Additionally, abrasion and deformation were considered as other types of surface defects as they are normally generated from the long-term rail-wheel contact and are visually identifiable, which are important properties for inspection and are the same as the other three surface defects mentioned earlier.

Following the categorisation of rail surface defects, existing inspection methods based on advanced sensory technologies were reviewed. These mainly include four different sensory technologies: 1) eddy current, 2) ultrasonic, 3) vision, and 4) laser. The basics of these sensory technologies were introduced to help understand their principles for rail surface defect inspection. For example, eddy current and ultrasonic are mainly based on the correlation between the defects and the reflected electric or ultrasonic signals. Vision mainly refers to HD cameras that capture the surface conditions of the rail through 2D images. Laser mainly refers to laser triangulation sensors that measure surface geometries based on the triangular reflection of structured laser lights. Since eddy current and ultrasonic technologies do not measure surface defects directly, they are naturally more sensitive to noise caused by the factors such as lift-off distances and the conductivity of the rail surface and mediums. These also limit their inspection speed. In comparison, laser and visual sensors can measure and characterise surface conditions directly, and they are truly non-contact. Therefore, vision- and laser-based sensory techniques were taken forward in this research through a detailed investigation of existing vision- and laser-based rail inspection approaches. The strengths and limitations of vision- and laser-based rail inspection as two individual techniques were further analysed. For example, the performance of vision-based systems is normally dependent on surface cleanliness and illumination equality, while laser-based systems are more robust to the complex surface and ambient lighting conditions. Similarly, vision-based inspection strategies are diverse, with numerous inspection algorithms available, while laser-based inspection methods show low diversity. These observations suggested the route to combine the strengths of different sensing techniques and eliminate their

shortcomings to improve their overall performance for rail inspection. This process is also known as multi-sensor technology. Accordingly, multi-sensor technology was introduced, with related applications in other fields reviewed. A multi-sensing system that integrates laser, vision, and other NDT sensors was then proposed, which aimed to improve the overall performance of laser and vision-based rail inspection.

The achievements of stage one can be concluded as follows:

- Identified dedicated rail surface defects in accordance with rail defect management standards, and clarified corresponding inspection requirements following related inspection standards.

- Identified fundamentals and characteristics of existing inspection solutions, and selected vision and laser as two candidate sensory technologies.

- Initially identified strengths and shortcomings of existing vision- and laser-based inspection systems and specific inspection algorithms.

- Developed a roadmap for combining the strengths of different sensors and eliminating their shortcomings to improve their overall performance for rail inspection.

7.2 Stage two – Technology implementation and integration

With the inspection targets and the candidate sensory technologies clarified, stage two of this research discussed two case studies, each of which applied one of the candidate sensors into specific rail inspection tasks. The first case study (Chapter 3) was around a vision-based system that was designed and deployed for switch rail condition monitoring and inspection. The second case study (Chapter 4) evaluated the performance of laser-based rail inspection by using a laser triangulation sensor as an onboard measuring unit. A system that uses the structure of most existing laser-based rail inspection systems was developed and tested in both laboratory and outdoor field environments. A new multi-sensing system that fully considered the complementarity between laser and vision sensors and the requirements for rail surface defect inspection was developed (Chapter 5) based on the literature and case studies.

Specifically, through the first case study, two main strengths of vision-based systems were identified: 1) they are easy to deploy and non-destructive in most cases. For example, the vision system in this case study was installed line-side, which is

approximately 3.0 m away from the switch rail. The status of the switch rail was monitored and inspected through high-resolution video frames captured by the system, which is non-destructive and has no interference to any existing equipment; 2) numerous vision-based detection algorithms that have been developed and optimised are available. These provide a lot of options for different inspection tasks. For example, multiple descriptors including SURF, CHT, and V-channel were tested to detect and locate the switch rail in the case study introduced in Chapter 3. Although vision-based inspection works are non-destructive, it was also observed that vision-based inspection algorithms normally rely on distinguishable surface texture features of the object against the background in the image. These features can be the object's intensity, colour, or shape, etc. However, it was found that there are many external factors in a practical application that can affect the surface texture features of the object in the image. Four main factors were discussed in the case study: 1) lighting conditions, 2) surface cleanliness, 3) adverse weather conditions, and 4) material degradation. To ensure the effectiveness of a vision-based inspection system, additional measures may need to be taken to mitigate the potential effects of these factors. For example, in the case study, reflective marker points and external infrared lighting sources were used and could ensure the effectiveness of the system in most cases. Additionally, combination with other sensory techniques, such as the laser as is introduced in the later chapters of this research, can be considered.

In the second case study, the self-designed laser-based inspection system was used to carry out rail inspection in the form of 2D rail profile measurement first, like most of the existing laser-based rail inspection methods. Defects can be detected by identifying specific geometric changes caused by them. However, two main shortcomings of 2D profile-based rail inspection were identified: (1) the shadowing effect of the laser line can limit the integrity of measured profiles. This brought problems; for example, some defects that are partially distributed behind steep edges cannot be accurately measured. Also, the integrity of measured profiles can influence the assessment of some defects that relies on the comparison between the measured and corresponding standard profiles, as the comparison process normally involves profile alignment for which the integrity of measured profiles determines the alignment accuracy; (2) the geometric information contained in 2D profiles is limited. This limits the diversity of inspection strategies. Also, most rail surface defects are longitudinal in nature, and thus cannot be

characterised and detected accurately in 2D. Accordingly, the second part of the case study introduced a cost-effective way of measuring the rail surface in 3D using a 2D laser triangulation sensor. The linear nature of the rail was utilised by aligning 2D profiles along the longitudinal direction of the rail, on which a 3D reconstruction method was developed to reconstruct the surface geometry of the rail from the sequence of 2D profiles. The testing results demonstrated that by developing the 3D reconstruction method, the surface conditions of the rail can be characterised more comprehensively. Three geometric feature descriptors (depth gradient, face normal, and face-normal gradient) were developed, which indicated the potential of rail inspection using laser-based 3D models. Compared to the vision-based system introduced in Chapter 3, the laser-based system was proven to be more accurate and showed more tolerance to some common external factors that have impacts on optical inspection systems. For example, since laser sensors measure the geometry of the rail directly, normal surface texture changes have less impact on laser measurement results. Similarly, since laser sensors actively project highly collimated laser light, they are more robust to external light fluctuations. However, some shortcomings of existing laser-based systems were also identified. For example, the operation of the laser sensor in the case study still relied on an additional mechanical frame to maintain its position and ascertain the measurement stability, which in return increased the system size and reduced its degrees of freedom. Also, compared to vision-based inspection, the inspection strategies for laser sensor data were still limited. These observations through the two case studies further identified the strengths and shortcomings of laser- and vision-based inspection approaches and clarified the complementarities between them.

The outcomes of the two case studies (Chapters 3 & 4) can be concluded as follows:

- For rail inspection, vision-based systems are easy to deploy, while laser-based systems are normally in large size which limits their flexibility of deployment to some extent.

- Vision-based systems place more emphasis on continuous or frequent monitoring, and their system accuracy and stability can be affected by various external factors which are, therefore, normally of secondary importance. In comparison, laser-based systems are more accurate in terms of characterising rail surface conditions.

- Vision-based systems have numerous inspection algorithms available, while laser-based inspection strategies are limited by the integrity and the dimension of measurement data and the few mature applications to refer to.

- 3D rail surface measurement based on a 2D laser sensor was proven to be feasible, which increases the dimension of measurement and can potentially broaden the laser-based inspection strategies.

- The works related to the vision-based inspection system in case study one and the laser-based 3D rail surface characterisation method in case study two have generated two published journal papers [10] [81].

Based on the two case studies in Chapters 3 and 4 and the relevant observations, a new multi-sensing system that fully considers the characteristics of laser and vision sensors for rail inspection was designed (Chapter 5). The new prototype used a 2D laser triangulation sensor as the main measuring unit considering its accuracy and stability, and integrated an IMU module to track its motion and orientation and a camera module to evaluate the validity of the measured frames. The development of the new prototype was also within the scope of the S-CODE project [11], which ensures its practicability and ascertains that the design satisfies industrial requirements. The final deployment fully used the camera module's strength of object identification and the IMU module's strength of motion tracking, yielding a portable multi-sensing rail inspection system with increased degrees of freedom and automation. With the multi-sensing data, an optimised profile-to-profile registration method was developed. The profiles captured from different angles can be registered to eventually provide a rail profile measurement with increased integrity. The system was configured for hand-held deployment within the S-CODE project. The testing results demonstrated that the system can measure the rail profile with a controllable registration error of under 0.05 mm, and the complete rail profile improves the accuracy of profile matching for wear measurement, with an accuracy in the order of 0.1 mm on average. Comparison with a commercial full-contact profiling system further demonstrated the performance of the proposed system in terms of both system accuracy, stability, speed, and greater measurement coverage.

In conclusion, the main contributions of this part of the work are as follows:

- A new prototype for a laser-based multi-sensing system was developed for rail inspection, which overcomes the dependence of the laser sensor on additional

mechanical frames, providing a more compact, autonomous, and portable prototype.

- From a measurement perspective, more information is provided in the multi-sensing dataset, allowing greater integrity of the rail profile to be achieved by accurately aligning multiple scans with profile-to-profile registration.

- From an inspection perspective, a complete rail profile improves the performance of profile matching for wear assessment in terms of both efficiency and accuracy.

- The works related to the new prototype have generated a journal paper and been published in *IEEE Transactions on Instrumentation and Measurement* [18].

7.3 Stage three – Intelligent inspection algorithms

The multi-sensing system proposed in stage two of this research demonstrated that an effective combination of laser and vision sensors contributes to the improvement of their overall performance as rail inspection instruments. More accurate and higher-dimensional multi-sensing data have also shown the potential to tackle another limitation of existing laser-based rail inspection, which is the limited availability of compatible inspection algorithms. For rail surface defect inspection, two fundamental problems are defect detection and classification. Existing surface defect detection and classification methods are mostly based on computer vision algorithms using 2D images. However, their practicability is often questioned as the imaging quality of vision-based systems can be easily affected by external factors as has been discussed in Chapter 3. In comparison, 3D laser measurement data have been proven to be more accurate in characterising rail surface conditions. Stage three of this research (Chapter 6) analysed the transferability of processing approaches between 2D images and 3D laser measurement data. The latest concepts of deep learning in computer vision were applied to processing 3D laser sensor data. Two new inspection strategies were developed: *1) end-to-end 3D pixel-level surface defect detection*, and *2) 3D model-based surface defect classification and evaluation*. These combine the strength of laser sensors in high-precision measurement with the strength of numerous vision algorithms. The promising results and the contributions of stage three of this research can be concluded as follows:

- A strategy allowing the application of 2D machine learning approaches to process 3D datasets was developed, to provide a more accurate and comprehensive rail surface characterisation method for surface defect inspection.

- An optimised deep semantic segmentation network was developed based on an FCN, which can process 3D laser sensor data directly and output pixel-level defect detection results in 3D. The trained model has been proven to be able to carry out end-to-end pixel-level surface defect detection with an average accuracy of up to 87.9%, which is comparable to or even better than most existing pixel-level defect detection methods that rely on high-quality 2D images. More importantly, it shows more tolerance to complex surface conditions such as rail surface contaminations and texture changes.

- Compared to existing defect detection methods using 2D image data, the proposed pixel-level defect defection method can not only locate the defect and describe its boundary accurately, but also allows for reconstruction of the rail and defect geometry in 3D, which provides essential references for further evaluation and maintenance works.

- With the detected defects in 3D, four geometric descriptors were extracted to form a low-dimensional pattern for each defect, which were then used to train multiple classifiers to further classify detected defects into three common classes and two severity levels. The classification accuracy of the trained classifiers can be up to 96.7%, which demonstrates the potential of the proposed strategy to automate defect classification and evaluation works that are conventionally human-dominated.

- The two inspection strategies introduced in stage three of this research further broaden the application scenarios of the proposed multi-sensing system, and also provide solutions for the development of multi-sensing inspection strategies in other relevant areas. Two journal papers have been generated on the work described in stage three of this research. One has been published in *IET Image Processing* [103] and another one is under review by IEEE which can refer to section 1.5 - 1 for details.

7.4 Conclusions

The research works in this thesis demonstrated the hypothesis that the overall performance of laser- and vision-based rail inspection can be improved by effectively combining the strengths of disparate sensors and eliminating their shortcomings. From the system point of view, a new prototype for a laser-based multi-sensing system was developed. Benefiting from the integration of a laser module, a camera module, and an IMU module, the proposed system is more autonomous and robust with increased degrees of freedom and a compact size. Dedicated data integration algorithms were developed to allow for higher-dimensional measurement and the solution to different inspection tasks. The system in combination with the optimised profile-to-profile registration method achieved 2D hand-held rail profile measurement with greater integrity. The system in combination with the specific 3D reconstruction pipeline achieved onboard 3D rail surface characterisation. The multi-sensing system showed better performance than existing systems of the same type that rely on a single sensor, in terms of measurement integrity, dimension, accuracy, and efficiency. From the inspection point of view, multi-sensing data contain comprehensive information and allow for the combination of different inspection algorithms. The latest concepts of machine learning and deep learning in computer vision were applied to processing high-precision 3D laser sensor data, yielding two new inspection strategies for the two fundamental surface defect inspection problems: 1) end-to-end 3D pixel-level surface defect detection, and 2) 3D model-based surface defect classification and evaluation. The performance of these two inspection strategies is promising in terms of both accuracy and efficiency. They can potentially replace or at least reduce the dependency on human inspectors to carry out refined rail inspection tasks more efficiently.

In summary, the proposed multi-sensing system and dedicated inspection algorithms form a systematic solution for rail measurement and surface defect inspection and demonstrate that the integration of multiple sensor technologies can provide improved inspection performance compared to those technologies being used independently. The work described in this thesis has formed five publications as listed in section 1.5.

7.5 Limitations and future work

Although the research work has demonstrated the hypothesis, the proposed multi-sensing rail inspection approach still has room for improvement. This can be summarised by the limitations from both system and inspection strategy points of view.

From the system point of view, the limitations that are worth considering in future work can be summarised in the following three points:

- As a multi-sensing system that uses a laser triangulation sensor as the main measuring unit, the performance of the laser sensor to a great extent determines the performance of the system. Therefore, those shortcomings of the laser sensor are the major concerns. The multi-sensing system in this research mitigates three shortcomings of existing laser-based rail inspection: 1) system size, 2) measurement integrity, and 3) the shadowing effect of the laser line. Another limitation of the laser triangulation sensor is the shadowing effect of the receiver as has been discussed in section 4.3. As this is an inherent shortcoming related to the imaging mechanism of laser triangulation sensors, it is out of the scope of this research. The shadowing effect of the receiver can mainly cause errors in the measurement of the depth of some defects with a large ratio of depth to width. In future work, additional sensors, such as a single-point laser sensor, can be considered to correct the depth measurement results of the defect. Additionally, a confidence index could also be used to indicate the maximum depth that the system can measure accurately for a defect with a specific width.

- The current system is only configured for hand-held deployment for 2D rail profile measurement and an onboard deployment for 3D railhead surface measurement. The sampling rate of the system after a trade-off to the accuracy was set to approximately 20 Hz, which is sufficient for the current deployments. In the future, interaction with other mobile actuators, such as a robotic arm or a drone, will be considered to explore more comprehensive and fully autonomous rail inspection solutions. In this configuration, it would also be appropriate to consider computational acceleration to support a greater rate of measurement.

- Currently, 3D measurement is only considered for the railhead surface where surface defects are mostly distributed. A complete 3D measurement that covers other parts of the rail should be considered in future work to increase the freedom of inspection and widen the system's application scenarios. This can potentially be achieved by aligning the complete 2D rail profiles in the longitudinal direction or using improved ICP to register 3D measurement results captured from different angles.

From the inspection point of view, the two main surface defect inspection strategies proposed in this research are based on supervised learning. Their performance can be improved considering the following three points:

- The current models are only trained and tested on 400 samples. Although data augmentation and transfer learning are applied to deal with the small dataset during the training process, inaccurate detections can occur when facing some special cases, as indicated by the experimental results in sections 6.2.5 and 6.3.3. A larger dataset with increased size and diversity can improve the model's generalisation, and thus is worth considering in future work.

- The networks used in this research are migrated from the existing architectures in computer vision. Although they have been optimised for the inspection task in this research, their structures are still complicated and normally contain millions of parameters. These involve massive computational costs, and thus a powerful PC was used as the computational platform. This, however, may not be accessible in other application scenarios such as field deployments. In the future, the structure of the networks and the training process can be continuously optimised to reduce the computational cost, to further improve its practicability.

- As supervised learning methods, all the training processes require predefined labels or categorisations. To manually label hundreds of samples is acceptable. However, for more complex inspection tasks, the number of samples may increase exponentially, which can pose challenges to the manual labelling process. Moreover, it is difficult to avoid human errors during the manual labelling process. Recently, semi-supervised and unsupervised learning methods learning have been a new trend in some areas such as reinforcement learning for autonomous driving, which are also considered to be a trend for inspection tasks in the future.

8 References

[1]   Department for Transport. *Transport Statistics Great Britain 2019*, 2019.

[2]   International Energy Agency. *The Future of Rail - Opportunities for Energy and the Environment*, 2019.

[3]   Rowshandel, H. *The Development of an Autonomous Robotic Inspection System to Detect and Characterise Rolling Contact Fatigue Cracks in Railway Track*, University of Birmingham, 2014.

[4]   Network Rail. *Network Rail's New Measurement Train*, 2014.

[5]   Cannon, D.F., K.O. Edel, S.L. Grassie, and K. Sawley. 'Rail defects: An overview,' in *Fatigue and Fracture of Engineering Materials and Structures*, 2003, **26**(10), pp. 865–887.

[6]   MERMEC. *Track Measuring Systems* [Full Rail Profile], 2017. Available from: http://www.mermecgroup.com.

[7]   EURAILSCOUT. *Switch Inspection & Measurement Fact Sheet*, 2014.

[8]   AREMA. *Development and Implementation of Automated Switch Inspection Vehicle*, 2011.

[9]   Rusu, M.F. *Automation of Railway Switch and Crossing Inspection*, University of Birmingham, 2017.

[10]   Ye, J., E. Stewart, and C. Roberts. 'Use of a 3D model to improve the performance of laser-based railway track inspection,' *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2018, **233**(3), pp. 337–355.

[11]   Roberts, C. *Switch and Crossing Optimal Design and Evaluation (S-CODE)*, 2020. Available from: http://www.s-code.info/.

[12]   U.S. Department of Transportation. Federal Railroad Administration. *Rolling Contact Fatigue: A Comprehensive Review*, 2011.

[13]   Hassankiadeh, S.J. *Failure Analysis of Railway Switches and Crossings for the Purpose of Preventive Maintenance*, KTH Royal Institute of Technology in Stockholm, 2011.

[14]   Capacity for Rail. *Operational Failure Modes of Switches and Crossings*, 2015, **CAPACITY4RAILSCP3-GA-2013-605650**.

[15]   Nissen, A., J. Colina, J.U. Eiby, et al. *Inspection of Switches & Crossings*, 2011: UIC.

[16]   Grassie, S.L. 'Rolling contact fatigue on the British railway system: Treatment,' *Wear*, 2005, **258**(7), pp. 1310–1318.

[17]   Royal Institute of Technology. *Failure Analysis of Railway Switches and Crossings for the Purpose of Preventive Maintenance*, 2011.

[[18]   Ye, J., E. Stewart, D. Zhang, et al. 'Integration of multiple sensors for noncontact rail profile measurement and inspection,' *IEEE Transactions on Instrumentation and Measurement*, 2021, **70**, pp. 1–12.

[19]   Papaelias, M., C. Roberts, and C.L. Davis. 'A review on non-destructive evaluation of rails: State-of-the-art and future development,' *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2008, **222**(4), pp. 367–384.

[20] Liu, S., Q. Wang, and Y. Luo. 'A review of applications of visual inspection technology based on image processing in the railway industry,' *Transportation Safety and Environment*, 2020, **1**(3), pp. 185–204.

[21] Alahakoon, S., Y.Q. Sun, M. Spiryagin, and C. Cole. 'Rail flaw detection technologies for safer, reliable transportation: A review,' *Journal of Dynamic Systems, Measurement, and Control*, 2017, **140**(2), 020801.

[22] Hodge, V.J., S. O'Keefe, M. Weeks, and A. Moulds. 'Wireless sensor networks for condition monitoring in the railway industry: A survey,' *IEEE Transactions on Intelligent Transportation Systems*, 2015, **16**(3), pp. 1088–1106.

[23] Kim, H.M., H.R. Yoo, and G.S. Park. 'Analysis of a defect signal deformations induced by eddy current in RFECT system for pipeline inspection,' *IEEE Transactions on Magnetics*, 2018, **54**(11), pp. 1–5.

[24] U.S. Department of Transportation. *Track Inspector Rail Defect Reference Manual*, 2015.

[25] Gonzalez, R.C. and R.E. Woods. *Digital Image Processing*, 2001: Addison-Wesley Longman Publishing Co., Inc.

[26] Micro-Epsilon. *scanCONTROL // 2D/3D Laser Scanner (Laser Profile Sensors)*, 2015.

[27] Gonzalez, R.C. and R.E. Woods. *Digital Image Processing*, 2018: Pearson.

[28] Li, Q. and S. Ren. 'A real-time visual inspection system for discrete surface defects of rail heads,' *IEEE Transactions on Instrumentation and Measurement*, 2012, **61**(8), pp. 2189–2199.

[29] Zhang, H., X. Jin, Q.M.J. Wu, et al. 'Automatic visual detection system of railway surface defects with curvature filter and improved Gaussian mixture model,' *IEEE Transactions on Instrumentation and Measurement*, 2018, **67**(7), pp. 1593–1608.

[30] Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, 1982: Henry Holt and Co., Inc.

[31] Wang, Y., H. Xia, X. Yuan, L. Li, and B. Sun. 'Distributed defect recognition on steel surfaces using an improved random forest algorithm with optimal multi-feature-set fusion,' *Multimedia Tools and Applications*, 2018, **77**(13), pp. 16741–16770.

[32] Zheng, H., L. Kong, and S. Nahavandi. 'Automatic inspection of metallic surface defects using genetic algorithms,' *Journal of Materials Processing Technology*, 2002, **125**, pp. 427–433.

[33] Choi, D.-C., Y.-J. Jeon, J.P. Yun, and S.W. Kim. 'Pinhole detection in steel slab images using Gabor filter and morphological features,' *Applied Optics*, 2011, **50**(26), pp. 5122–5129.

[34] Bodnarova, A., M. Bennamoun, and S.J. Latham. 'Textile flaw detection using optimal Gabor filters,' in *Proceedings of the 15th International Conference on Pattern Recognition. ICPR-2000*, 2000.

[35] Luo, Q., X. Fang, L. Liu, C. Yang, and Y. Sun. 'Automated visual defect detection for flat steel surface: A survey,' *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(3), pp. 626–644.

[36] Bishop, C.M. *Pattern Recognition and Machine Learning (Information Science and Statistics*, 2006: Springer-Verlag.

[37] Xie, X. 'A review of recent advances in surface defect detection using texture analysis techniques,' *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, 2008, **7**(3), 1–22.

[38] Wang, J., Q. Li, J. Gan, H. Yu, and X. Yang. 'Surface defect detection via entity sparsity pursuit with intrinsic priors,' *IEEE Transactions on Industrial Informatics*, 2020, **16**(1), pp. 141–150.

[39] Yun, J.P., D. Kim, K. Kim, et al. 'Vision-based surface defect inspection for thick steel plates,' *Optical Engineering*, 2017, **56**(5), 053108.

[40] Silvén, O., M. Niskanen, and H. Kauppinen. 'Wood inspection with non-supervised clustering,' *Machine Vision and Applications*, 2003, **13**(5), pp. 275–285.

[41] Jeon, Y.-J., D.-C. Choi, J.P. Yun, and S.W. Kim. 'Detection of periodic defects using dual-light switching lighting method on the surface of thick plates,' *ISIJ International*, 2015, **55**(9), pp. 1942–1949.

[42] Cross, G.R. and A.K. Jain. 'Markov random field texture models,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1983, **PAMI-5**(1), pp. 25–39.

[43] Ayed, I.B., N. Hennane, and A. Mitiche. 'Unsupervised variational image segmentation/classification using a Weibull observation model,' *IEEE Transactions on Image Processing*, 2006, **15**(11), pp. 3431–3439.

[44] Network Rail. *Rail Failure Handbook*, 2001.

[45] Xiong, Z., Q. Li, Q. Mao, and Q. Zou. 'A 3D laser profiling system for rail surface defect detection,' *Sensors (Basel)*, 2017, **17**(8), 1791.

[46] Xue-Wu, Z., D. Yan-Qiong, L. Yan-Yun, S. Ai-Ye, and L. Rui-Yu. 'A vision inspection system for the surface defects of strongly reflected metal based on multi-class SVM,' *Expert Systems with Applications*, 2011, **38**(5), pp. 5930–5939.

[47] LeCun, Y., Y. Bengio, and G. Hinton. 'Deep learning,' *Nature*, 2015, **521**(7553), pp. 436–444.

[48] Chollet, F. *Deep Learning with Python*, 2017: Manning Publications Co.

[49] Natarajan, V., T. Hung, S. Vaikundam, and L. Chia. 'Convolutional networks for voting-based anomaly classification in metal surface inspection,' in *2017 IEEE International Conference on Industrial Technology (ICIT)*, 2017.

[50] Liu, B., J.C. Brigham, H.E. Jun, X. Yuan, and H.U. Huiling. 'Roll contact fatigue defect recognition using computer vision and deep convolutional neural networks with transfer learning,' *Engineering Research Express*, 2019, **1**(2), 025018.

[51] He, Y., K. Song, Q. Meng, and Y. Yan. 'An end-to-end steel surface defect detection approach via fusing multiple hierarchical features,' *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(4), pp. 1493–1504.

[52] Redmon, J., S. Divvala, R. Girshick, and A. Farhadi. 'You Only Look Once: Unified, real-time object detection,' in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[53] Shelhamer, E., J. Long, and T. Darrell. 'Fully convolutional networks for semantic segmentation,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(4), pp. 640–651.

[54] Yang, F., L. Zhang, S. Yu, et al. 'Feature pyramid and hierarchical boosting network for pavement crack detection,' *IEEE Transactions on Intelligent Transportation Systems*, 2020, **21**(4), pp. 1525–1535.

[55] Dong, H., K. Song, Y. He, et al. 'PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection,' *IEEE Transactions on Industrial Informatics*, 2019, **16**(12), pp. 7448–7458.

[56] Badrinarayanan, V., A. Kendall, and R. Cipolla. 'SegNet: A deep convolutional encoder-decoder architecture for image segmentation,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(12), pp. 2481–2495.

[57] Song, K. and Y. Yan. 'A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects,' *Applied Surface Science*, 2013, **285**, pp. 858–864.

[58] Huang, Y., C. Qiu, Y. Guo, X. Wang, and K. Yuan. 'Surface defect saliency of magnetic tile,' in *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, 2018.

[59] Micro-Epsilon. *Operating Instructions scanCONTROL 29xx/BL*, 2017.

[60] Liu, Z., J. Sun, H. Wang, and G. Zhang. 'Simple and fast rail wear measurement method based on structured light,' *Optics and Lasers in Engineering*, 2011, **49**(11), pp. 1343–1351.

[61] Giri, P. and S. Kharkovsky. 'Detection of surface crack in concrete using measurement technique with laser displacement sensor,' *IEEE Transactions on Instrumentation and Measurement*, 2016, **65**(8), pp. 1951–1953.

[62] Loccioni. *Felix: Robot for Switches & Crossings Inspections* [10-2021]. Available from: https://www.loccioni.com/en/railway/.

[63] Yang, Y., L. Liu, B. Yi, and F. Chen. 'An accurate and fast method to inspect rail wear based on revised global registration,' *IEEE Access*, 2018, **6**, pp. 57267–57278.

[64] Molleda, J., R. Usamentiaga, A.F. Millara, et al. 'A profile measurement system for rail quality assessment during manufacturing,' *IEEE Transactions on Industry Applications*, 2016, **52**(3), pp. 2684–2692.

[65] Shelburne, E. 'Producing high-quality 3-D models for specimen research with the Creaform 3-D Scanner at the Sternberg Museum of Natural History,' *Biodiversity Information Science and Standards*, 2018, **2**, e25908.

[66] Creaform. *Understanding Portable 3D Scanning Technologies*, 2020: Creaform, Canada.

[67] Chen, P., Y. Hu, W.-T. Li, and P.-J. Wang. 'Rail wear inspection based on computer-aided design model and point cloud data,' *Advances in Mechanical Engineering*, 2018, **10**(12).

[68] Besl, P.J. and N.D. McKay. 'A method for registration of 3-D shapes,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, **14**(2), pp. 239–256.

[69] Liu, H., Y. Li, Z. Ma, and C. Wang. 'Recognition and calibration of rail profile under affine-distortion-based point set mapping,' *IEEE Transactions on Instrumentation and Measurement*, 2017, **66**(1), pp. 131–140.

[70] Yi, B., Y. Yang, Q. Yi, W. Dai, and X. Li. 'Novel method for rail wear inspection based on the sparse iterative closest point method,' *Measurement Science and Technology*, 2017, **28**(12), 125201.

[71]   Bentley, J.L. 'Multidimensional binary search trees used for associative searching,' *Communications of the ACM*, 1975, **18**(9), pp. 509–517.

[72]   Yang, Y., L. Liu, B. Yi, and F. Chen. 'Dynamic inspection of a rail profile under affine distortion based on the reweighted-scaling iterative closest point method,' *Measurement Science and Technology*, 2019, **30**(11), 115202.

[73]   Li, Y., X. Zhong, Z. Ma, and H. Liu. 'The outlier and integrity detection of rail profile based on profile registration,' *IEEE Transactions on Intelligent Transportation Systems*, 2019, **21**(3), pp. 1074–1085.

[74]   Zhou, P., K. Xu, and D. Wang. 'Rail profile measurement based on line-structured light vision,' *IEEE Access*, 2018, **6**, pp. 16423–16431.

[75]   Dobmann, G., J.H. Kurz, A. Taffe, and D. Streicher. 'Development of automated non-destructive evaluation (NDE) systems for reinforced concrete structures and other applications,' in *Non-Destructive Evaluation of Reinforced Concrete Structures*, C. Maierhofer, H.-W. Reinhardt, and G. Dobmann (Eds.), 2010: Woodhead Publishing, pp. 30–62.

[76]   Li, R., J. Liu, L. Zhang, and Y. Hang. 'LIDAR/MEMS IMU integrated navigation (SLAM) method for a small UAV in indoor environments,' in *2014 DGON Inertial Sensors and Systems (ISS)*, 2014.

[77]   Chen, M., S. Yang, X. Yi, and D. Wu. 'Real-time 3D mapping using a 2D laser scanner and IMU-aided visual SLAM,' in *2017 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, 2017.

[78]   Zhang, J. and S. Singh. 'LOAM: Lidar Odometry and Mapping in Real-time,' in *Robotics: Science and Systems*, 2014.

[79]   National Rail. *Inspection and Maintenance of Permanent Way*, 2018.

[80]   Abtuslimited. *ABT4670 - Switch & Crossing Confirmation Gauge,* 2021; Available from: http://www.abtus.com/en/products/.

[81]   Ye, J., E. Stewart, Q. Chen, L. Chen, and C. Roberts. 'A vision-based method for line-side switch rail condition monitoring and inspection', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2021, **0(0)**, pp. 1-11.

[82]   Bay, H., T. Tuytelaars, and L. Van Gool. 'SURF: Speeded Up Robust Features'. in *Computer Vision – ECCV 2006*. 2006. Berlin, Heidelberg: Springer Berlin Heidelberg.

[83]   Su, Y., W. Gao, Z. Liu, S. Sun, and Y. Fu. 'Hybrid Marker-Based Object Tracking Using Kinect v2', *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(9), pp. 6436-6445.

[84]   Rekimoto, J. and Y. Ayatsuka. 'CyberCode: designing augmented reality environments with visual tags', in Proceedings of DARE 2000 on Designing augmented reality environments. 2000, *Association for Computing Machinery: Elsinore*, Denmark. pp. 1–10.

[85]   Haralick, R.M., S.R. Sternberg, and X. Zhuang. 'Image Analysis Using Mathematical Morphology', *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1987, **PAMI-9**(4), pp. 532-550.

[86]   Canny, J.: 'A Computational Approach to Edge Detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, **PAMI-8**(6), pp. 679-698.

[87]   MICRO-EPSILON. *Instruction Manual scanCONTROL 29xx/BL*, 2008.

[88]   Tam, G.K.L., Z. Cheng, Y. Lai, et al.: 'Registration of 3D Point Clouds and Meshes: A Survey from Rigid to Nonrigid', *IEEE Transactions on Visualization and Computer Graphics*, 2013, **19**(7), pp. 1199-1217.

[89]   MathWorks. '*pcdownsample'* in *MathWorks*, 2015.

[90]   Marton, Z.C., R.B. Rusu, and M. Beetz. 'On fast surface reconstruction methods for large and noisy point clouds'. in *2009 IEEE International Conference on Robotics and Automation*. 2009.

[91]   Zhang, D., E. Stewart, J. Ye, M. Entezami, and C. Roberts. 'Roller bearing degradation assessment based on a deep MLP convolution neural network considering outlier regions', *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(6), pp. 2996–3004.

[92]   Ash, C., M. Dubec, K. Donne, and T. Bashford. 'Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods', *Lasers in medical science*, 2017, **32**(8), pp. 1909-1918.

[93]   Analog Devices. *Triaxial Inertial Sensor with Magnetometer-ADIS16405*, 2019.

[94]   Analog Devices. *Tilt Measurement Using a Dual Axis Accelerometer*, 2012.

[95]   Otsu, N. 'A threshold selection method from gray-level histograms', *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, **9**(1), pp. 62–66.

[96]   Su, C. and R.M. Haralick. 'Recursive erosion, dilation, opening, and closing transforms', *IEEE Transactions on Image Processing*, 1995, **4**(3), pp. 335-345.

[97]   Lam, L., S. Lee, and C.Y. Suen. 'Thinning methodologies-a comprehensive survey', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, **14**(9), pp. 869-885.

[98]   Atherton, T.J. and D.J. Kerbyson. 'Size invariant circle detection', *Image and Vision Computing*, 1999, **17**(11), pp. 795-803.

[99]   Fan, H., P.C. Cosman, Y. Hou, and B. Li. 'High-Speed Railway Fastener Detection Based on a Line Local Binary Pattern', *IEEE Signal Processing Letters*, 2018, **25**(6), pp. 788-792.

[100]   Fan, J., J. Yang, D. Ai, et al.: 'Convex hull indexed Gaussian mixture model (CH-GMM) for 3D point set registration', *Pattern Recognition*, 2016, **59**, pp. 126-141.

[101]   British Steel. *Flat-Bottomed Vignole Rail*, 2018.

[102]   Greenwood Engineering. *MiniProf Digital Profile Measuring System*, 2015.

[103]   Ye, J., E. Stewart, D. Zhang, Q. Chen, and C. Roberts. 'Method for automatic railway track surface defect classification and evaluation using a laser-based 3D model', *IET Image Processing*, 2020, **14**(12), pp. 2701-2710.

[104]   Ronneberger, O., P. Fischer, and T. Brox. 'U-Net: Convolutional Networks for Biomedical Image Segmentation'. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015. Cham: Springer International Publishing.

[105]   Russakovsky, O., J. Deng, H. Su, et al. 'ImageNet large scale visual recognition challenge', *International Journal of Computer Vision*, 2015, **115**(3), pp. 211–252.

[106] Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia. 'Pyramid Scene Parsing Network'. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.

[107] Keshet, L. *Math 103*, 2012: University of British Columbia.

[108] Zhou, Z.H. and S. Liu. *Machine Learning*, 2021: Springer Singapore.

[109] Berrar, D. 'Cross-validation', in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, et al. (Eds), 2019: Academic Press, Oxford, pp. 542–545.

[110] Ye, J., E. Stewart, and C. Roberts. 'Use of a 3D model to improve the performance of laser-based railway track inspection', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2018, **233**(3), pp. 337–355.