A C V P R

Zheng Liu
Hiroyuki Ukida
Pradeep Ramuhalli
Kurt Niel *Editors*

Integrated Imaging and Vision Techniques for Industrial Inspection

Advances and Applications

Springer

Advances in Computer Vision and Pattern Recognition

More information about this series at http://www.springer.com/series/4205

Zheng Liu · Hiroyuki Ukida
Pradeep Ramuhalli · Kurt Niel
Editors

Integrated Imaging and Vision Techniques for Industrial Inspection

Advances and Applications

Springer

Editors
Zheng Liu
University of British Columbia
Kelowna, BC
Canada

Hiroyuki Ukida
University of Tokushima
Tokushima
Japan

Pradeep Ramuhalli
Pacific Northwest National Laboratory
Richland, WA
USA

Kurt Niel
University of Applied Sciences Upper
 Austria
Wels
Austria

Preface

Machine vision is defined as a process of integrating a variety of technologies and computational methods to provide imaging-based information. The scope of machine vision is broad, and it has been applied to a wide range of applications including robot guidance, quality assurance, sorting, material handling, optical gauging, automated assembly, industrial inspection. This book focuses on the industrial inspection, which is to ensure the safety and reliability of the products through the measurements and tests of certain characteristics for an object or activity. Integrated imaging and machine vision techniques are the key technologies to achieve such measurements and tests. With the evolution of machine vision hardware, such as smart camera, LED illumination, time-of-flight camera, multicore processor, and graphics processing unit (GPU), it becomes possible to offer a cost-effective solution for inspection tasks with high-performance capabilities. Moreover, the expansion of wavelength gains the visibility of features not available in visible spectrum and thus greatly enhances the reliability and performance.

Advanced machine vision systems may incorporate multiple imaging and/or vision modalities to provide robust solutions to complex situations and problems in industrial applications. A diverse range of industries, including aerospace, automotive, electronics, pharmaceutical, biomedical, semiconductor, and food/beverage, and manufacturing, have benefited from recent advances in multimodal inspection technologies. This book is a collection of pioneering contributions in machine vision from academia and industries for varied inspection tasks. It highlights both the advances in technologies and vision system integration for practical applications. The advances provide an insight into recent progresses and developments of imaging and vision techniques for varied industrial inspection tasks, while the applications present the state of the art of imaging and vision system integration, implementation, and optimization.

This book consists of two major parts, that is, (Part I) Technology Advances and (Part II) Applications and System Integration for Vision-Based Inspection. The "Technology Advances" provide an insight into recent progresses, developments, and future trends of imaging and vision techniques for varied industrial inspection

tasks, while the "applications and system integration" present the state of the art of imaging and vision system integration, implementation, and optimization. The first chapter is contributed by the editorial team. A comprehensive review of the technology advances in machine vision is presented. This chapter screens the state of the art in machine vision technologies in the light of hardware, software tools, and evolution of algorithms for industrial inspection. The inspection techniques beyond visual spectrum, such as infrared thermal imaging, ultraviolet imaging, acoustic and ultrasonic imaging, gamma-ray and X-ray imaging, Terahertz imaging, millimeter-wave imaging, are described. This chapter also highlights multispectral and multimodal imaging in inspection, which can provide complementary information not available in unimodal imaging systems.

Part I of this book contains eight chapters. Ibarra-Castanedo et al. describe the reflectography/transmittography in the near-infrared (NIR) and short-wave infrared (SWIR) spectra and the infrared thermography in the mid-wave infrared (MWIR) and long-wave (LWIR) infrared bands. The NIR/SWIR reflectography/transmittography is applied to inspect semitransparent composite materials such as glass, aramid–phenolic, and a variety of natural fibers, while MWIR/LWIR thermography can detect surface and subsurface defects in varied types of materials. The complementary nature between NIR/SWIR reflectography/transmittography and MWIR/LWIR thermography makes it possible to provide an integrated solution for the specific inspection task. Metal surface introduces challenges to the visual inspection with its varying surface characteristics from one material to another. In the chapter contributed by Huber-Mörk et al., different configurations or setups for image acquisition and illumination are presented for the applications such as rail surface inspection, weld quality assessment, inspection of holograms on metallic foils, and reconstruction of coins. The algorithms for weld quality assessments are also described and discussed in detail. Multiview approaches are proposed for the inspection of optically variable security elements on metallic foil, and light-field data processing is applied to the inspection of coins. In automated visual inspection, the purpose of registering multiple images is to associate corresponding pixels to the same physical points on the object under inspection. In the next chapter, Penz et al. describe a real-time nonparametric 2D image registration method, namely "FlexWarp," which is based on image warping with grids and image pyramids. The FlexWarp algorithm can automatically select and assign control points, where only the most salient image features are correlated. A parallel implementation of the time-critical parts with GPU is applied to the print inspection in real time. Further discussions on the settings and performance of FlexWarp are available in this chapter. The background and recent advances in the development of portable optical coordinate measurement machine (CMM) and handheld scanner are explained in the chapter written by Larue et al. The unique capabilities of the new portable 3D measurement offer a more reliable and operator-independent tool for 3D metrology. More specifically, the dynamic referencing, automatic drift detection, and automatic alignment functionality make the measurement productive and efficient at a low cost. The significant impact to the 3D metrology is foreseen. A solution for fast 3D shape inspection is presented in the chapter by Ishii. A light-section method, which uses

multisided mirror between an image sensor and object under inspection, is proposed for 3D shape measurement. The measurement system is implemented with a high-speed vision hardware, which allows real-time image (1024 × 1024) processing at 1000 fps. Printed circuit board (PCB) with solders was inspected to verify the measurement accuracy at submillimeter level. The European project "3DComplete" is to create a low-cost and flexible quality inspection system capable of capturing 2.5D color data for the completeness of the inspection. The chapter by Munaro et al. is devoted to providing details of this work. The 3DComplete system employs a single color camera and a laser triangulation system to capture 3D data, which are then combined into a color 2.5D model in real time. This 2.5D color scanner can be applied to the inspection of objects within a production line for the completeness. This system is competent to perform the inspection tasks that remain challenges for state-of-the-art 2D-based methods. X-ray computed tomography (XCT) is used to detect and characterize the heterogeneities of materials by Kastner et al. Their work on cone beam XCT with micro-focus and sub-micro-focus X-ray sources is reported in the next chapter. Advanced image processing and visualization methods are developed for quantitative evaluation of the properties of metals and polymeric materials. To tackle the challenges introduced by curved surface with highly specular reflection to the defect inspection, Sho et al. proposed a novel vision system in their chapter, in which a set of images are captured under various illumination directions. This image set is then integrated to construct a single image suitable for defect detection, which is based on morphologic and template matching operations.

Part II of this book mainly focuses on applications and system integration. There are six chapters in this part. In modern manufacturing, industry robots are employed to conduct the inspection tasks with machine vision systems. The chapter from Eitzinger et al. covers topics relevant to the implementation of robotic inspection systems for industrial quality control from machine vision's point of view. Progresses in increasing the level of autonomy of the inspection process and making the inspection system easier to handle are highlighted. Condition assessment is critical for the management of civil infrastructure. Among all the methods, visual inspection is a key technology for assessing the physical and functional conditions of civil infrastructure. Koch et al. contribute their chapter to the machine vision techniques for the condition assessment of civil infrastructure. The chapter begins with the current practices in civil infrastructure condition assessments. The state-of-the-art machine vision techniques available for the condition assessment of civil infrastructure are described. The benefits and limitations of each technique as well as the corresponding challenges associated with the techniques are highlighted. Several case studies on inspecting bridges, buildings, and roads are presented to demonstrate the effectiveness of these techniques in condition assessment. Another work from the "3DComplete" project is presented in the chapter by Carlon et al., which targets inspecting a complex solid part. A 3D vision system is mounted on a manipulator robot arm to perform the inspection. This system integrates three software modules, i.e., visual inspection, 3D simulation, and motion planning of the manipulator. The system can automatically generate the needed points of view in order to perform 3D reconstruction and automatic visual inspection. Next chapter

contributed by Crawford et al. describes the ultrasonic techniques for the inspection of nuclear power plant welded components. Modern ultrasonic imaging techniques such as synthetic aperture focusing technique (SAFT), phased-array (PA) technology, and sound field mapping were applied to assess the integrity of the component and estimate the remaining useful life through the data interpretation and analyses. The use of electromagnetic wave imaging for concrete and wood structures is reported by Fujii et al. in Chap. 14, which comprises the description for the principle, system architecture, and image processing approaches. The last chapter of the book about magneto-optic imaging (MOI) is contributed by Deng et al. MOI, which combines electromagnetic and optical energies, is applied to the inspection of surface and subsurface cracks and corrosion in aircraft structures. The generated image is easy to interpret in comparison with conventional eddy current impedance data. The relatively new technique is insensitive to lift-off variations and can cover large inspection area. This chapter describes state of the art in the recent MOI system development and advances in processing algorithms.

This book offers the in-depth description of advances in inspection methodologies and machine vision technologies for specific needs, which makes it an excellent reference for researchers on developing innovative solutions to tackle practical challenges. The engineers will also benefit from the applications at both the system and component levels and be able to assure a smooth and efficient manufacturing process with a successful inspection system. For dealing with real engineering stuff, some thoughts about pitfalls and barriers of such systems are given for better understanding of their behavior in the automation process in the industry. This book will bridge the gap between theoretical knowledge and engineering practice for university graduate students, who are studying imaging, machine vision, and industrial inspection.

This book is a dedicated team work of all the contributors. Their great research work achieves the excellence of this collection. They deserve the greatest appreciations for their significant contributions. During the course of this project, the editorial team received continuous support from the publisher. We are grateful for all their help and effort on this work. Finally, Mr. Chenxi Yang at Toyota Technological Institute (Japan) is acknowledged for his sincere help on creating the index page of this book.

Contents

Chapter 1
Industrial Inspection with Open Eyes: Advance with Machine Vision Technology

Zheng Liu, Hiroyuki Ukida, Kurt Niel and Pradeep Ramuhalli

Abstract Machine vision systems have evolved significantly with the technology advances to tackle the challenges from modern manufacturing industry. A wide range of industrial inspection applications for quality control are benefiting from visual information captured by different types of cameras variously configured in a machine vision system. This chapter screens the state of the art in machine vision technologies in the light of hardware, software tools, and major algorithm advances for industrial inspection. The inspection beyond visual spectrum offers a significant complementary to the visual inspection. The combination with multiple technologies makes it possible for the inspection to achieve a better performance and efficiency in varied applications. The diversity of the applications demonstrates the great potential of machine vision systems for industry.

Contents

Z. Liu (✉)
University of British Columbia Okanagan, Kelowna, British Columbia, Canada
e-mail: zheng.liu@ieee.org

H. Ukida
Tokushima University, Tokushida, Japan
e-mail: ukida@tokushima-u.ac.jp

K. Niel
University of Applied Sciences Upper Austria (Campus Wels), Wels, Austria
e-mail: kurt.niel@fh-wels.at

P. Ramuhalli
Pacific Northwest National Laboratory, Richland, WA, USA
e-mail: pradeep.ramuhalli@pnnl.gov

1.1   Introduction

Visual inspection is to identify whether a particular attribute is present or properly located in a predetermined area [1]. As described in [2], inspection is to determine whether an object deviates from a given set of specifications. Machine vision systems are designed to release human operators in industrial inspection process and achieve a more robust and high quality of performance of manufacturing process and quality control. Different from computer vision, which refers to a broad terms of the capture and automation of image analysis, machine vision needs an engineering to system design with additional hardware I/O and computer networks to transmit information [3]. One of the most common applications of machine vision is the inspection of products, such as integrated circuits, vehicle parts and components, food and pharmaceuticals [4]. An early survey reported automated visual inspection systems and techniques from 1988 to 1993 [2]. Both the benefits and challenges on using of CAD data or models, which contain exact specifications of an object, in inspection are highlighted and discussed. Based on the inspected features, the inspection tasks can be categorized into four basic groups: dimensional characteristics, surface characteristics, structural quality, and operational quality [5]. A variety of software and hardware solutions for the machine vision system development are reviewed in the year 2003 [5]. In the review [6], real-time performance and verification of industrial machine vision systems as well as the temporal reliability were discussed and described.

The flexibility and complexity of modern manufacturing process bring challenges for the assurance of speed of production and quality of product. The increasing demands on constant precision and reliability need sophisticated machine vision systems to take significantly varied inspection tasks. Machine vision systems are evolving with the technology advances, such as imaging sensors, digital interfaces, illumination, computational capability, artificial intelligence, communication, and network. The emerging of low-cost and high-efficient embedded vision systems and smart cameras makes it possible to build a

scalable machine vision system for varied industrial applications. The integration of multiple cameras and multimodal imaging systems offers a more robust solution for the difficulties in industrial inspection. Fusing the visual information and those beyond visual spectrum can achieve a comprehensive inspection with less uncertainty. Industrial inspection is benefiting from such advances for improved accuracy and performance. This purpose of this chapter is to screen the advances in industrial inspection and review the technologies, which promote such changes and progresses.

This chapter is organized as follows. The first Sect. 1.2 is devoted to the description of general architecture of machine vision systems. The advances of hardware, software, and algorithms for machine vision systems are highlighted in Sect. 1.3. Vision systems beyond visual spectrum are briefly described in the next Sect. 1.4. The expansion in the applications of machine vision is presented in Sect. 1.5. This chapter is summarized in the final Sect. 1.6.

1.2 System Architecture for Modern Machine Vision System

Modern machine vision system consists of digital input/output devices and computer networks for automatic operation of equipments or quality control systems. There are basically three categories: PC-based vision system, embedded vision system, and smart cameras. Figure 1.1 shows the three types of machine vision systems. The smart camera-based systems cover a wider range of applications, while embedded vision and PC-based systems offer more flexibility and higher performance due to the increased complexity of the overall system.

A typical PC-based machine vision system consists of the following components as illustrated in Fig. 1.2 [7]:

- Cameras and optics
- Lighting

Fig. 1.1 Overview of three typical machine vision systems [133]
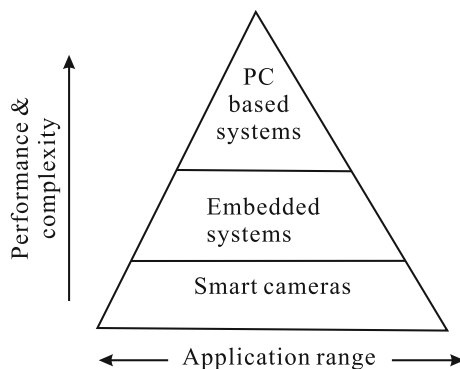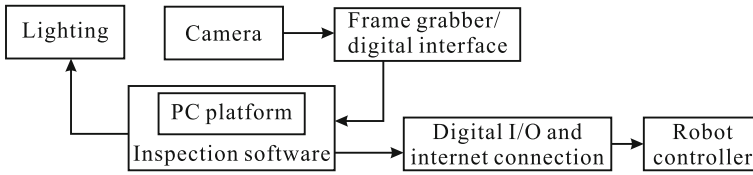
Fig. 1.2 A typical PC-based vision system for industrial inspection and quality control

- Frame grabber
- PC platform
- Inspection software
- Digital I/O and network connection

The PC-based machine vision system employs one or more cameras and lenses to capture a picture of the object under inspection [7]. Different types of cameras can be used, such as monochrome camera, RGB color camera, and progressive-scan or line-scan camera. To assure a better quality of the image, the object is illuminated with the lighting device. The lighting of high-frequency fluorescent, light-emitting diode (LED), incandescent, and quart-halogen fiber optic can be configured in various shapes and sizes with a variety of intensities [7]. Frame grabbers or video capture cards provide low-level interface capabilities with other system components and host computer [5]. It can also control the camera by setting the triggering, exposure/integration time, shutter speed, etc. The PC platform runs the inspection software to process acquired image data and even make a binary decision, e.g., accept/reject. For varied inspection tasks, algorithms need to be tailored with proper software tools. The data exchange and communication with outside systems and databases are done through the digital I/O interface and/or network connection. The PC-based machine vision system should be configured based on the specific requirements and goal of the inspection task.

Embedded vision is the merging of computer vision and embedded systems. The low-cost, powerful, and energy-efficient processors makes it possible to incorporate vision capabilities into a wide range of embedded systems for using of visual inputs [8, 9]. Embedded vision becomes a key technology for the automated inspection in manufacturing and quality control systems [8]. Embedded vision is usually implemented with a combination of processing elements, such as CPU, high-performance digital signal processor (DSP) processor, and highly parallel engine [10]. The popular types of processors used in embedded vision system include high-performance embedded CPU, graphics processing unit (GPU) + CPU, DSP + accelerators + CPU, field-programmable gate array (FPGA) + CPU, and application-specific standard product (ASSP) + CPU [10].

As described in [9], embedded vision systems are facing a number of challenges and constrains in the compute resources. The issues that should be considered during the embedded vision system development include the following:

- Specialized computing hardware, such as DSP, GPU, and FPGA;
- Smaller memory footprint for space limitations;
- Bandwidth issues for real-time analysis;
- Security issues during communication and data transferring; and
- Algorithm optimization for embedded vision system.

To develop real-time embedded vision systems, both general-purpose and vender-specific tools can be employed, for example, the open-source computer vision library OpenCV, which needs to be ported to non-PC platforms for embedded implementation. Embedded vision has been widely adopted in the industrial inspection applications and is proliferating broad markets with the emergence of high-performance, low-cost, energy-efficient programmable processors [10, 11].

Smart cameras integrate lenses, sensors, processors, camera-to-computer interfaces, and software in one camera system. As described by Wolf et al. in [12], smart cameras output processed image data with high-level descriptions of a scene and perform real-time analysis of what they see. The overall architecture of a smart camera is illustrated in Fig. 1.3. The application-specific processing (ASP) is performed by embedded algorithms on a per channel basis [13, 14].

Shi and Real [15] categorized smart cameras into the following classes:

- Integrated smart cameras item,

  - Smart camera on a chip or single-chip smart cameras,
  - Embedded smart cameras,
  - Stand-alone smart cameras item,

- Compact-system smart cameras, and
- Distributed smart cameras.

From top to the bottom of the list, the level of integration between ASP and image capture parts of the camera system decreases, while the flexibility and complexity of the camera system increase [15]. Generally, in vision industry, conventional PC-based vision systems are being replaced by smart cameras due to its relatively low cost, simplicity, integration, and reliability [14, 16]. The smart camera-based systems can carry out real-time processing operations within the camera at high speeds and lower cost [14]. In [17], an investigation was conducted to compare three smart camera architectures for real-time machine vision system. Processing speed and power consumption appear to be the most important metrics. Vision-systems-on-chip is the trend for the future [18, 19].
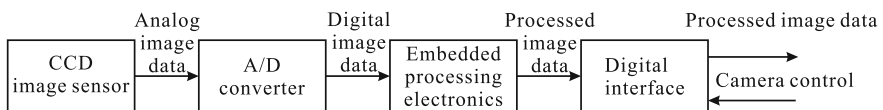
Fig. 1.3 Architecture of smart camera [14]

1.3   Advances of Machine Vision Systems

1.3.1   Advances in Hardware

Referring to the machine vision system architecture in Fig. 1.2, four major parts of machine vision hardware, i.e., illumination and lighting, imaging sensors and cameras, digital camera interfaces, and computation units, are reviewed in this section. A brief summation is illustrated in Fig. 1.4.

1.3.1.1   Illumination and Lighting

Illumination and lighting is critical to the success of machine vision applications. A light source produces light, which is reflected from the object under inspection. A camera then collects the reflected light for further analysis. The emitted wavelengths from the light source should be compatible with the color of the object. There are six commonly used light sources in machine vision as listed in Table 1.1 [20]. The choice of the light source depends on the types of inspection tasks, scale of inspection area, requirements on wavelength and brightness. The selection criteria include life expectancy, cost-effectiveness, heat output, intensity, stability, and flexibility [20]. For instance, LED lighting is replacing fluorescent and quartz halogen lighting due to its stability, long lifetimes, cost-effectiveness, and flexibility. By properly configuring the LED illuminators of different wavelengths, the lighting source is designed and controlled to meet the inspection requirements.
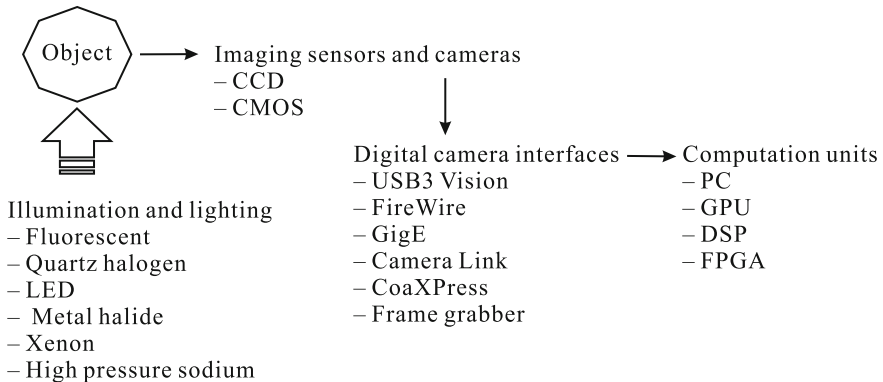
Fig. 1.4 Summary of the hardware in machine vision systems

Table 1.1 Commonly used light sources in machine vision [20]

Fluorescent	Metal halide (mercury)
Quartz halogen—fiber optics	Xenon
LED—light-emitting diode	High-pressure sodium

For example, an integrated illumination system with LED array is described in [21]. This system includes a reflector to reflect light from LED array toward the illumination area and is suitable for use with area-array-based scanning equipments like video camera [21].

When multiple LED lights of different colors are used or the illumination intensity of a single LED needs to be adjusted, the control of LED light is essential to achieve an efficient machine vision processing. The use of LED light controller can thus optimize illumination intensity and precision trigger sequencing between cameras and lights. A controller can conduct the following operations [22]:

- Pulse or strobe control: synchronize the switch-on time of lighting with the exposure time of camera;
- Overdriving control: increase the intensity of LED light for a short, defined period of time;
- Power supply control: provide a highly stable current supply for LED lighting;
- Multilighting schemes control: implement intensity control and high-speed synchronization with single or multiple triggers for multiple lighting configurations;
- Remote configuration control: set lighting system parameters remotely.

In high-speed applications, LED controllers can strobe the light at high speeds and narrow pulse widths with external trigger inputs. Moreover, LED illuminators can be overdriven to provide high illumination in a short duration of time for a higher contrast image. For more flexibility, FPGA can be used in the remote control of reconfigurable LED illumination system [23].

Machine vision lighting should maximize feature contrast while minimize the contrast elsewhere [24]. According to [24], the following rules of thumb should be considered when choose lighting for a machine vision system:

- Use bright light to detect missing material;
- Use appropriate wavelength for accurate component placement;
- Use non-diffused light to detect cracks in glass;
- Use diffused light to inspect transparent packaging;
- Use color to create contrast;
- Use strobed light for rapidly moving parts;
- Use infrared light to eliminate reflections;
- Use infrared light to diminish color variation.

The lighting system should be capable of producing and modulating light, and communicating lighting command signals with the controller in the machine vision system [25].

1.3.1.2 Imaging Sensors and Cameras

Machine vision systems use cameras that employ different types of image sensors [26]. There are two main types of image sensors, i.e., charge-coupled device

(CCD) and complementary metal-oxide semiconductor (CMOS). According to the description in [27], in a CCD sensor, every pixel's charge is transferred through a very limited number of output nodes to be converted to voltage, buffered, and sent off-chip as an analog signal. In contrast, each pixel in a CMOS sensor has its own charge-to-voltage conversion, and the sensor also includes amplifiers, noise correction, and digitization circuits. Thus, the chip outputs digital bits. CCDs are previously given more preference due to its higher quality in comparison with CMOS sensors [28]. Technical challenges and progresses in CCD are described in [29]. The CCD process and design technology can be improved to meet the application requirements. However, CMOS sensors have evolved rapidly for better quality and performance. A view of relative advantages of these two techniques is given in Table 1.2.

X-ray CCD image sensors are to replace film and analog video imaging techniques with a number of benefits, including increased operation efficiency, digital image processing capability, lower X-ray dose with real-time image capability, adoption of archival infrastructure, data share, and communication capabilities [30]. According to [30], there are four primary X-ray technologies currently being used:

- Computed radiography,
- Image-intensified CCD (II-CCD) camera,
- Flat direct detector, and
- Flat indirect detector.

Some conventional technologies do not support real-time operations or suffer from the presence of residual image information in successive images, which is known as "image lag." II-CCD cameras have significant limitations in terms of physical size, inability to support challenging projections, vulnerability, and shorter intervals between calibrations [30]. To overcome these drawbacks, wafer-scale CMOS image sensor was developed with comparatively low noise and need for lower X-ray dose [31]. The negligible lag (less than 0.1 %) and high frame-rates make it suitable for use in large panels intended for dynamic X-ray applications [31]. CMOS image detectors offer numerous advantages including the ability to record smaller image details at higher resolutions [30].

Indium Gallium Arsenide (InGaAs) detectors are highly sensitive to energy in the near-infrared (NIR) and shortwave-infrared (SWIR) wavebands from 900 to 1700 nm [32]. Thus, InGaAs focal plane array (FPA) is used in lightweight sensor head. InGaAs FPA sensors are typically designed for use in night vision and thermal inspection applications [33].

Table 1.2 Comparison of CCD and CMOS image sensor [28]

CCD	CMOS
· High image quality	· High speed
· Light sensitivity	· On-chip system integration
· High fill factor and low noise	· Low manufacturing cost
· Excellent global shutter	· High level of CMOS process innovation

High-sensitive camera can acquire images of dark objects which is not possible for ordinary cameras [34]. According to [34], high-sensitivity cameras are grouped into charge-integration and electron-multiplication types based on their operating methods. The charge-integration-type camera achieves enhanced sensitivity from increasing the available light intensity to boost signal strength by extending the charge-integration time (exposure time). The electron-multiplication-type cameras gain the enhanced sensitivity from boosting the signal strength by some kind of electron-multiplication method. The cameras belonging to each type are given in Fig. 1.5. There is great potential to use high-sensitive camera in industrial inspection, where only illumination at low-light levels is available.

High-definition video camera was used to collect field data of a concrete bridge on a mobile carrier [35]. An image stitching software was then used to generate a single high-resolution composite image by combining the image frames acquired at varied angles and distances. The composite image was then processed to determine the structures current condition as related to crack size, location, and distribution.

Time-of-flight (ToF) cameras measure the time of flight of modulated light, which is reflected from the object of interest. Such distance measurement can capture a 3D image. For instance, the light can be modulated with an RF carrier and phase shift of the carrier bounded back from object is measured to quantify the distance [16]. In [16], a prototype ToF system using an off-the-shelf, high-resolution, hight-speed camera is described, which can capture VGA images at video rates. More details of the system are available in the patent document [36].

1.3.1.3 Digital Camera Interfaces

The interface to camera is one of the most important choices when setting up a vision system [37]. The common interfaces include capture boards (frame grabber), USB, FireWire, GigE, Camera Link, and CoaXPress. Capture boards are also known as frame grabber, which are used to capture image frames from an analog video signal or a digital video stream. Capture board also refers to PCI cards that are
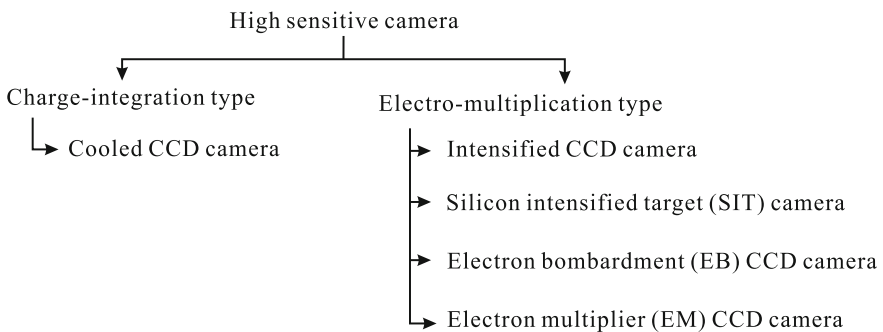
Fig. 1.5 High-sensitive camera types [34]

necessary to acquire and interpret the data from digital camera interfaces, but are not based on standard computer connectors [38].

Universal serial bus (USB) is an industry standard, which defines the cables, connectors, and protocols for connection, communication, and power supply between computers and electronic devices. Digital camera is one of such devices. As cameras with an USB 2.0 interface follow a simple design concept and come with little memory, data loss happens frequently [37]. In contrast, USB 3.0 provides a high-speed easy-to-use interface without frame grabber, and the potential to avoid data loss. The USB3 Vision standard, which was released in 2013 based on USB 3.0, defines necessary elements to make USB 3.0 compliant with industry vision applications. It uses nine pins in the cable and a unicast dual simplex data interface to allow bidirectional data transmission. USB3 Vision standard is being widely used in general-purpose industrial imaging and machine vision applications (Table 1.3).

	Advantages	Limitations
USB 2.0	• Common interface on every PC • Low costs • Nonstandard driver	• CPU load must be optimized
USB 3.0/USB3 Vision	• Standard interface • Low CPU load • Energy management with low power consumption and suspend modes • One cable solution • Variable image size • Low latency and jitter times	• Cable lengths • Challenges with multicamera setups

FireWire standard defines a stable method for exchanging data between cameras and computer. Also known as IEEE 1394, the standard evolves from FireWire-a to

Table 1.3 Comparison of camera interface [37]

Interface	Cable lengths	Bandwidth	Real-time trigger	Power over cable
USB 2.0	5 m or more	60 MB/s gross 24 MB/s net	No	Yes
USB 3.0	5 m	Up to 625 MB/s gross 400–500 MB/s net	No	Yes
FireWire	up to 4.5 m	FireWire-(a) 50 MB/s gross 37.5 MB/s net FireWire-(b) 100 MB/s gross 75 MB/s net max 64 MB/s per device	No	Yes
GigE	up to 100 m passive	125 MB/s gross 100 MB/s net	No	Yes
Camera link	up to 10 m for 85 MHz	base 255 MB/s medium 510 MB/s full 680 MB/s	Yes	Yes

FireWire-b interface with a double increases in the bandwidth. However, it only supports up to 64 MB/s (IEEE 1394b). The cable length can reach up to only 4.5 m and the flexibility of a FireWire cabling is limited.

Advantages	Limitations
· Include trigger ready signaling, integration enabled signaling, and change on the fly optimization	· Interface not available on every PC
· Very low CPU load	· Limited cable lengths
· Real-time capability	· Limited bandwidth

Gigabit Ethernet (GigE) becomes a popular interface for high-performance industrial cameras. The adoption of Gigabit Ethernet communication protocol brings benefits like long cable length and high data transmission rate to machine vision systems [39]. In 2011, version 2.0 of GigE standard, which includes non-streaming device control, was released [39]. The new version allows transmission of compressed images (JPEG, JPEG 2000, and H.264), accurate synchronization of multicamera systems and enhanced support for multitap sensors.

Advantages	Limitations
· Standard interface on nearly all PC hardware	· CPU load must be optimized
· High data transmission rates	
· Easy infrastructure setup for multicamera use	
· Possible long cable lengths	
· One cable solution possible (PoE-power over Ethernet)	

Camera Link is a hardware specification that standardizes the connection between cameras and frame grabbers. It defines a complete interface which includes provisions for data transfer, camera timing, serial communications, and real-time signaling to the camera [40]. Camera Link built for real-time and high-bandwidth parallel communication. Camera Link has become the standard interface for line-scan cameras based on the high line rates and the large amount of data that these cameras can generate [37]. Camera Link is currently in version 2.0. Future development will be the Camera Link HS [40].

Advantages	Limitations
· High bandwidth for data streams	· Needs special frame grabber
· Proven standard for easy product interoperability	· High-end peripheral
· Possible one cable solution, i.e., power over Camera Link (PoCL)	· Limitation of cable length

CoaXPress (CXP) is high-speed, point-to-point, serial communication standard for the transmission of video [41]. It interfaces the connection between cameras and frame grabbers with a standard 75-Ω coaxial cable. The combination of standard coaxial cable and high speed makes CoaXPress the next-generation high-speed camera interface. It allows for up to 6.25 Gb/s transmission rates with cable lengths up to 100 m [38]. Multiple cables can be used for speeds of up to 25 Gb/s. Like POE, Power-over-Coax (PoC) is an available option. However, a CoaXPress frame grabber is required. CoaXPress has the following features [41]:

- High data rates,
- Long cable lengths,
- Real-time behavior through fixed, low latency,
- Precise triggering capability,
- Flexible and reliable through the use of standard coax,
- Ease of integration: video, communication, control, and power over a single cable,
- Cost-effective cable solution,
- Hot pluggable, and
- Plug-n-play.

When multiple camera systems are considered, GigE Vision and USB3 Vision are better candidates in comparison with Camera Link as they are based on Internet protocol standards and each camera is addressable. CoaXPress has advantages in multicamera applications [42].

1.3.1.4   Computation Units

GPU, a processor with massive parallel architectures and tremendous memory bandwidth, makes it suitable for accelerated image processing [43]. Varied GPUs offer different levels of performance for the computational needs. A GPU can be used together with a CPU to accelerate machine vision applications, such as the space flight applications [44] and product line inspection [45]. Compute Unified Device Architecture (CUDA) is a proprietary compiler and toolkit for programming NVIDIA GPUs. Meanwhile, a GPU can also be programmed using industry standard software frameworks, i.e., DirectX and OpenCL [43]. open computing language (OpenCL) provides a framework for programming different processors, such as CPUs, GPUs, DSPs, FPGAs, and other processors, across varied platforms. There are also other libraries with image processing specific functions available for GPU programming such as GUPImage on iOS framework [46] and Matrox Imaging Library (MIL) [43].

Digital signal processors (DSPs) are microprocessors specialized for signal processing algorithms and applications [8]. Thus, DSPs are more efficient than general-purpose CPUs for kinds of machine vision applications. To ensure sufficient performance, DSPs are often supplemented with one or more coprocessors [8].

A typical DSP chip for vision applications consists of a CPU, a DSP, and multiple coprocessors [8].

FPGAs (field-programmable gate arrays) are flexible logic chips that can be reconfigured at the gate and block levels. The implementation is through the selection of I/O interfaces and on-chip peripherals tailored to the specific applications [8]. The configuration of the FPGA's hardware is adaptive to image sensors with various interfaces [47]. Thus, FPGAs provide a hardware design flexibility with high performance and low power consumption. A low-cost machine vision system, which was based on FPGA, was developed to detect the position of the compact disk in CD/DVD duplicators [48]. Operations such as image selection, restoration, resizing, and segmentation were implemented. More sophisticated algorithms for image processing can be implemented with FPGA as well [49]. According to the benchmark study [50], FPGAs are better for low-level algorithms, which can be run in parallel, such as filter kernels, while high-level algorithms with complex branches or control loops and operate on data widths which are a multiple of 8 bits are preferred for implementing on DSPs.

1.3.2   Evolution of Algorithms

The general machine vision algorithm consists of a sequence of stages to process acquired images as illustrated in Fig. 1.6. The initial stages are committed to correcting geometric distortion introduced by imperfect lenses, enhancing contrast, and stabilizing images [8]. The second set of stages convert raw images into information about objects by identifying objects' edges, motion, color, size or other attributes. The final set of stages make inferences about objects by classification.

Figure 1.7 shows image processing algorithms categorized by Dechow [51] for varied inspection tasks, which include characterization of features and defects, and 2D and 3D guidance. Image features are highlighted and extracted through operations, such as filtering, contrast enhancement, threshold, and segmentation. Blob analysis measures the parameters of connected features of discrete objects in an image [51]. Correlation and geometric search algorithms find a predefined feature, which is presented in a new image with different sizes and scales, as well as some geometric transformations. The template matching is performed to classify and interpret images. More sophisticated approaches such as neural networks and support vector machines (SVMs) can be employed to conduct the classification tasks.
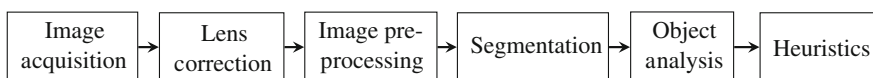
Fig. 1.6 A typical vision algorithm pipeline [8]
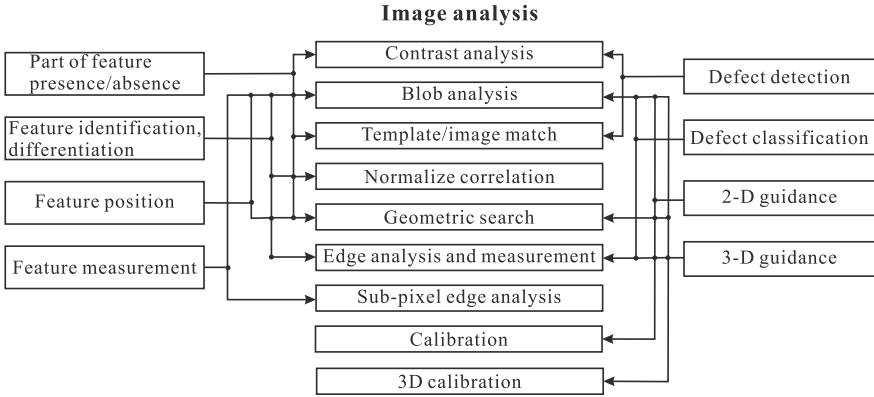
Image analysis

Fig. 1.7 Algorithms for varied inspection tasks [51]

For defect detection, the algorithms in Fig. 1.7 are configured to work together to conduct the analysis. Contrast analysis and template matching can be used to detect defects on a part. To further classify the defects, blob or edge analysis can be applied to characterize the defects and compare with known parameters. Sub-pixel resolution can be achieved by measuring the position of the line, point, or edge in the image with an accuracy exceeding the nominal pixel resolution of that image [51].

1.3.2.1 Algorithms for Image Processing and Analysis

The major machine vision algorithms are described in [52] and summarized in Fig. 1.8. These algorithms need to be configured and integrated for specific inspection applications. Details are available in [52]. This subsection will discuss the algorithms that are not covered by the Fig. 1.8. More details of these algorithms can be found in monographs [52–54].

The reconstruction of 3D surface can facilitate the quantitative analysis of surface topography. In industrial inspection, 3D reconstruction may detail the surface deformation, cracks, flaws, and roughness, which provide concrete evidence for quality control in the manufacturing process. There are various approaches that have been proposed to capture 3D surface information.

In [55], a light sectioning method is applied to the inspection of raw steel blocks. A laser light stripe is projected on the steel block and a camera is used to capture the laser light profile from a different direction. With the object moving in one direction, a complete 3D data set is acquired. The 3D reconstruction algorithm employs a geometric transformation to eliminate vibrations caused by movement. A spline interpolation and singular value decomposition methods are proposed to retrieve the depth map of the steel blocks. The obtained result can further be classified to identify the specific type of surface flaws [56].
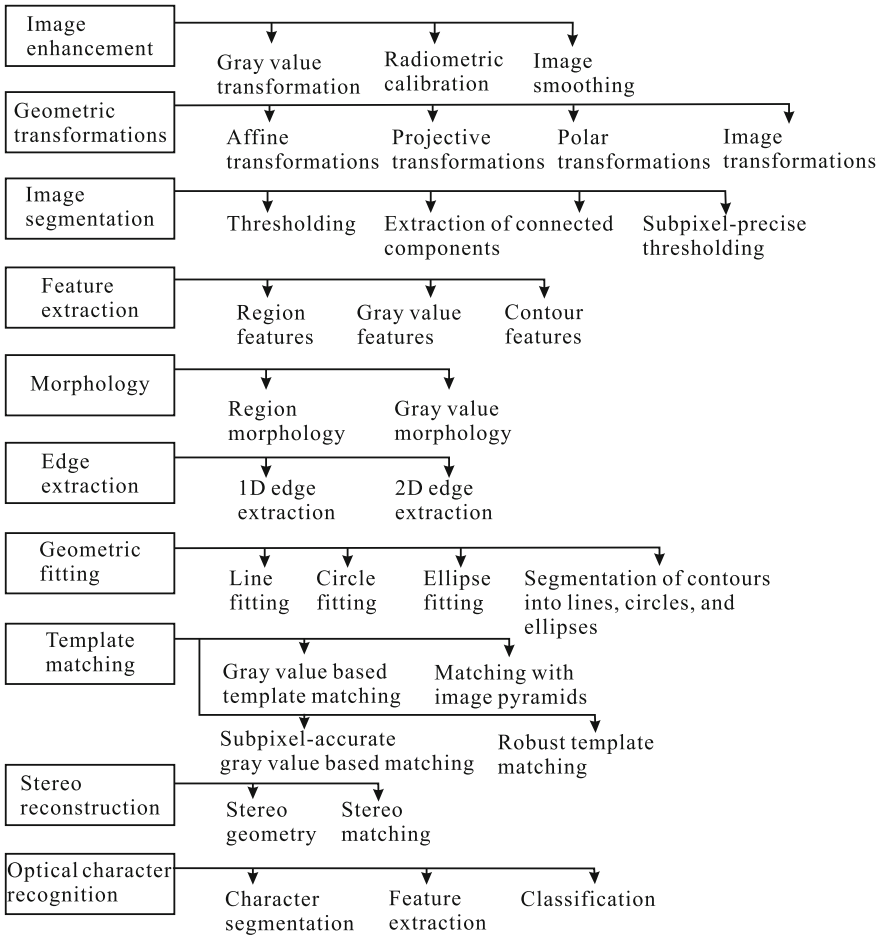
Fig. 1.8 Summary of machine vision algorithms [52]

In [57], surface reconstruction from the line-scanned image of a vision system, namely "edge of light (EOL)," is described. The EOL system targets the detection of surface deformation in aircraft lap joints due to hidden corrosion. The use of edge light can magnify the tiny chance on the surface. With a calibration by laser displacement sensor, the EOL image can be reconstructed to quantify the surface deformation [58]. The potential level of corrosion can be estimated.

Object detection or classification usually depends on the discriminative features from images. The features are based on the certain statistic or computational models, which rely on human experts. In [59], evolution-constructed (ECO) features are proposed for object detection. According to [60], feature construction is a process that discovers missing information about the relationships between features and augments the space of features by inferring or creating

additional features. The ECO features are automatically constructed by employing a standard genetic algorithm to discover series of transforms that are highly discriminative. ECO features have found great potential in recognizing objects from image.

Convolutional neural networks (CNN), which consist of multiple layers of small neuron collections, are inspired by biological processes [61]. CNN is a type of feed-forward neural network where the individual neurons are tiled in such a way that they respond to overlapping regions in the visual field [62]. Generally, a convolutional neural network includes different types of layers, e.g., convolutional layer, rectified linear unites (ReLU) layer, pooling layer, and fully connected layer. The max-pooling layers combine the outputs of neighboring neurons. If the outputs are integrated into an image, pooling layers will reduce its resolution. Various combinations of convolutional layers and fully connected layers are applied at the end of or after each layer [63].

CNN was used to classify a large set of high-resolution images in [64]. As illustrated in Fig. 1.9, the CNN contains eight learned layers, i.e., five convolutional and three fully connected layers. Response-normalization layers follow the first and second convolutional layers. Max-pooling layers follow both response-normalization layers as well as the fifth convolutional layer. The output of the last fully connected layer is fed to a 1000-way softmax which produces a distribution over the 1000 class labels. The network was trained with stochastic gradient descent on two NVIDIA GPUs for about one week. With the extracted latent features from the image, CNN has shown great potential for the applications of image classification and recognition.

1.3.2.2 Software Tools and Libraries

Both commercial and open-source software tools are available for machine vision applications. Table 1.4 lists some commercial tools for rapid prototyping of vision system for industrial inspection. Some packages offer intuitive and interactive interfaces for configuration and build of the algorithms. In addition, there are open-source software tools or libraries for vision applications.
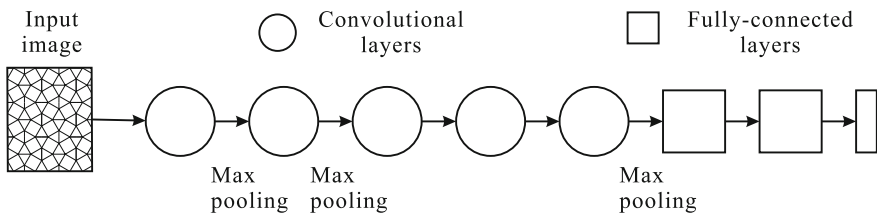
Fig. 1.9 An example architecture of convolutional neural network for image classification [64]

Table 1.4 Commercial software packages and tools

Vendor	Software package and tool	Description	References
Matrox	Matrox imaging library (MIL)	C library for C/C++ applications, including functions for image capture, processing, analysis, annotation, display, and archiving, and tools for calibrating, enhancing and transforming images, locating objects, extracting and measuring features, reading character strings, and decoding and verifying identification marks	[134]
National instrument	NI Vision Builder	Configurable software tool with built-in deployment interface for building, bench marking, and deploying machine vision applications, and with the ability to set up complex decisions to control digital I/O devices and communicate with peripherals	[135]
MVTec	HALCON	A general-purpose machine vision software with a comprehensive vision library covering all kinds of industrial applications	[136]
	MERLIC	A all-in-one software product with an image-centered user interface and intuitive interaction for quickly building machine vision applications without programming	[137]
Tordivel	SCORPION	An independent and open software tool with configurable 3D framework for industrial machine vision	[138]
Teledyne DALSA	Sapera	A suite of image processing and analysis functions with over 400 image processing primitives, bar code tool, pattern matching tools both area-based and edge-based, OCR, color, blob analysis, measurement and calibration tools for perspective and lens correction	[139]

The representative open-source software includes the following:

- OpenCV [65],
- SimpleCV [66],
- Visualization toolkit (VTK) [67],
- Darwin [68], and
- ROVIS [69].

OpenCV means open-source computer vision, which is a cross-platform C/C++ library of vision functions [70]. OpenCV provides an infrastructure to build real-time vision applications. It contains over five hundred functions that cover many areas in computer vision, such as factory product inspection, medical imaging, security, user interface, camera calibration, stereo vision, and robotics [65]. It also

contains a general-purpose machine learning library for statistical pattern recognition and clustering. OpenCV is written in optimized C and if further optimization is preferred, Intel's integrated performance primitive (IPP) libraries are recommended. The IPP libraries implement low-level optimization in many varied algorithms [65].

SimpleCV stands for simple computer vision and provides an easy-to-use Python framework integrating open-source computer vision libraries and algorithms [71]. It can interface with Webcams, Kinects, FireWire, IP cameras, and even mobile phones. Written in Python, SimpleCV inherits the easy-to-use nature of the language and makes the programming process easy and smooth.

The Visualization Toolkit (VTK) is an open-source C++ library for 3D graphics, modeling, image processing, volume rendering, and visualization [67, 72, 73]. The interpreted application layer is implemented in Tcl/Tk [73]. VTK is a cross-platform application, which enables high-quality and robust code.

DARWIN is an open-source platform-independent C++ framework for machine learning and computer vision research [68]. DARWIN provides stable, robust, and efficient libraries as described below:

1. *Base library* [drwnBase]—core infrastructure and utility routines for the whole framework;
2. *IO library* [drwnIO]—common input/output functionalities;
3. *Machine learning library* [drwnML]—implementations of various algorithms for supervised and unsupervised machine learning and utilities for dimensionality reduction, unconstrained optimization, linear programming, and maintaining sufficient statistics and collating classification results;
4. *Probabilistic graphical models* [drwnPGM]—functionality for inference in structured probability spaces over discrete random variables; and
5. *Computer vision library* [drwnVision]—based on OpenCV library and additional functionality for scene understanding.

DRWN provides infrastructure for data management, logging, and configuration; a consistent interface to standard machine learning and probabilistic graphical models algorithms; and well documented and easy-to-extend source code.

The ROVIS is another open-source software developed by the ROVIS research group, which integrates machine vision algorithms and techniques for the applications of scene and environment understanding [69]. According to [69], it provides a complete implementation of an object recognition and 3D reconstruction chain, e.g., camera calibration, image acquisition, filtering and segmentation, object recognition, and 3D scene reconstruction.

1.3.2.3 Automated Algorithm Performance Characterization

For software tools, libraries, or packages used in machine vision, there are a variety of algorithms that perform similar tasks. Thus, there is a need for efficient means to assess algorithm performance for specific task so that the "best" algorithm can

match the application automatically. The algorithms' performance may vary with the image contents. In [74], an automated performance characterization tool was demonstrated. Such tools should enable [74] the following:

- Automatically selecting the best candidate from multiple algorithms for a specific problem;
- Automatically identifying the best parameters for an algorithm and a particular image; and
- Introducing state-of-the-art algorithms to the application.

The performance analysis tools consider two probability values, i.e., probability of detection (P_d) and probability of false alarm (P_{fa}), which are defined as follows:

$$P_d = \frac{\text{Number of True Positive}}{\text{Number of Ground Truth Detections}} \tag{1}$$

$$P_{fa} = \frac{\text{Number of Detections} - \text{Number of True Positives}}{\text{Number of Detections}} \tag{2}$$

Thus, the automated algorithm characterization is to define a performance metric, such as the curve of P_d against P_{fa}, and determine which algorithm to use given a collection of images in terms of the optimal performance. The tools should also be able to predict the behavior on new data sets. The automated performance characterization tools facilitate the rapid development of machine vision applications and will be available as a Web-based software service in the future [74].

1.4 Vision Beyond Visual Spectrum

All the possible electromagnetic radiation consists of the electromagnetic spectrum as shown in Fig. 1.10 and corresponding wavelengths are listed in Table 1.5. The wavelength of the visible light ranges approximately from 390 to 770 nm. After the

The whole electromagnetic spectrum
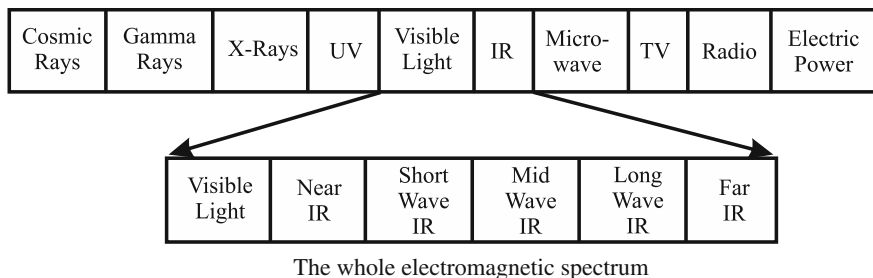
Fig. 1.10 The electromagnetic spectrum and infrared segments

Table 1.5 The list of electromagnetic wavelength

Electromagnetic wave	Wavelength λ (µm)
Cosmic rays	$\lambda < 10^{-7}$
Gamma rays	$10^{-4} > \lambda > 10^{-8}$
X-rays	$0.1 > \lambda > 10^{-7}$
UV	$0.39 > \lambda > 0.01$
Visible light	$0.77 > \lambda > 0.39$
IR	$10^3 > \lambda > 0.77$
Microwave	$10^6 > \lambda > 10^3$
TV and radio wave	$10^{11} > \lambda > 10^6$
Electric power	$\lambda > 10^{10}$

visible light comes the infrared (IR), which ranges from 770 to 1 mm and is further divided into five segments, e.g., near IR, shortwave IR, mid-wave IR, long-wave IR, and far IR.

Infrared radiation has the following characteristics [75]:

- Invisible to human eyes;
- Small energy;
- Long wavelength; and
- Emitted from all kinds of objects.

1.4.1 Infrared Thermal Imaging

Objects having temperature more than 0 K (-273.15°) can generally emit infrared radiation across a spectrum of wavelengths. The intensity of an object's emitted IR energy is proportion to its temperature. The emitted energy measured as the target's emissivity, which is the ratio between the emitted energy and the incident energy, indicates an object's temperature. At any given temperature and wavelength, there is a maximum amount of radiation that any surface can emit. If a surface emits this maximum amount of radiation, it is known as a blackbody. Planck's law for blackbody defines the radiation as [76] follows:

$$I_{\lambda,b}(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1} \tag{3}$$

where $I(\lambda, T)$ is the spectral radiance or energy per unit time, surface area, solid angle, and wavelength (Uit: $Wm^2µm^{-1}sr^{-1}$). The meaning of each symbol in above equation is listed below [77]:

λ wavelength (m)

T Temperature (K)

h Planck's constant (J/Hz)

c speed of light (m/s)
k Boltzmann's constant (J/K)

Usually, objects are not blackbodies. According to Kirchhoff's law, there is $R + \varepsilon = 1$, where ε is the emissivity and R is the reflectivity. Emissivity is used to quantify the energy-emitting characteristics of different materials and surfaces. The emitted energy of an object reaches the IR sensor and is converted into an electrical signal. This signal can be further converted into a temperature value based on the sensor's calibration equation and the object's emissivity. The signal can be displayed and presented to the end users. Thus, thermography can "see in the night" without an infrared illumination. A typical system to detect infrared radiation is usually configured as shown in Fig. 1.11 [75]. The amount of radiation increases with temperature; therefore, the variations in temperature can be identified by thermal imaging. The IR cameras can generally be categorized into two types: cooled infrared detectors and uncooled infrared detectors. They can detect the difference in infrared radiation with insufficient illumination or even in total darkness. The use of thermal vision techniques can be found in numerous applications such as military, law enforcement, surveillance, navigation, security, and wildlife observation [78]. The IR image can provide an enhanced spectral range that is imperceptible to human beings and contribute to the contrast between objects of high-temperature variance and the environment. Compared with a visual image, the IR image is represented with a different intensity map. The same scene exhibits different features existing in different electromagnetic spectrum bands.

The infrared detectors include two types, i.e., thermal type and quantum type [75, 79]. The thermal type has no wavelength dependence, while the quantum type is wavelength-dependent. Thermal detectors do not require cooling, but the response time is slow and detection capability is low. Quantum detectors own a higher detection performance and a faster response time cycle, but they have to be cooled for accurate measurement except those used in the near-infrared region.

Infrared light is more effective at penetrating certain materials than visible wavelengths [80]. Near infrared (wavelength of 850 and 940 nm) is often used in machine vision, which is available in large-die/high-output LED emitters [80]. The use of infrared light can discreet and covert illumination, and minimize light pollution. However, IR light cannot be used with color cameras.

As reported in [35], a high-end infrared camera was used to locate possible delamination and spalling of concrete through the monitoring of temperature

| Infrared source | → | Transmission system | → | Optical system | → | Detector | → | Signal processing |
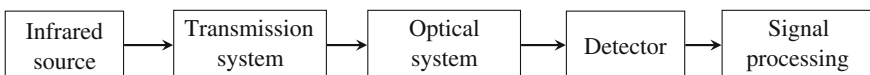
Fig. 1.11 A typical configuration for the detection of infrared radiation [75]

variations on the concrete surface of high-way bridges. Statistical and analytical relationship between characteristics of temperature variation and inherent damage of the concrete was established, which resulted in an automatic damage classification system that could classify the damage rate into three categories.

One major field of thermal imaging is food and agriculture industry—thanks to its non-contact, noninvasive, and efficiency nature. The applications include crop water stress monitoring, disease and pathogen detection, ice nucleation monitoring, and determination of fruit yield. [81]. In addition, post-harvest operations, i.e., quality assessment, also employ thermal imaging techniques. Reference [81] presents a review of such applications.

Infrared thermography (IRT) also shows its great potential in condition monitoring as temperature is a common indicator of structural health [82]. The non-intrusive nature of the technique makes it suitable for a wide variety of applications including monitoring of civil structures, electrical and electronic components, weld, corrosion, and deformation. IRT has been widely applied for assessing various components in nuclear industries, such as the monitoring of the annulus fan bearings and motor faults located on top of the radioactive liquid waste tank, cooling water pump, slurry pump, exhaust fans, and electrical components such as panels, circuit breakers, and relays. [82]. In aerospace industry, several types of composite structures, such as carbon and glass fiber-reinforced polymers, honeycomb structures, sandwich panels, are inspected with IRT for delamination, de-bonds, foreign inclusions, porosity, water ingress, node failure, and core crushing [82].

1.4.2   Ultraviolet (UV)

Like shorter and longer infrared lighting, shorter ultraviolet(UV) lighting is also useful for specific applications. UV wavelengths are used to exit the material under inspection and the material fluoresces in visible spectrum [83]. Thus, UV inspection is conducted in the visible region with UV filters by a monochrome or color camera. Color cameras can identify the object of interest through different wavelengths or colors. Meanwhile, high-current LEDs in UV can provide a focused high-brightness illumination at greater working distances [83].

There are two distinct techniques in UV imaging, i.e., reflected-UV imaging and UV-fluorescence imaging. Reflected-UV imaging uses ultraviolet light to illuminate a surface and capture the reflected or scattered light in UV band [84]. Reflected-UV imaging is often used to detect surface scratches and contamination. In contrast, the UV-fluorescence imaging detects the signal in visible or infrared band.

In [85], gunfire spectral signatures were characterized in the UV spectrum to detect the muzzle flash of an AK-47 assault refile. Ultraviolet fluorescence was used to detect freeze-damaged oranges based on the peel oil constituents on the surface [86]. Freeze-damaged oranges exhibit a fine pattern of 12-mm bright yellow dots on the peel when viewed under long-wave UV light. With an automated machine

vision system, moderate or severe freeze conditions can be identified. It is also possible to detect rotten mandarin oranges with the fluorescent substance using 365-nm excitation wavelength as reported in [87].

1.4.3 Acoustic and Ultrasonic Imaging

Internal packaging defects can cause field failures of an electrical board. Acoustic micro imaging (AMI) offers a non-intrusive imaging of object internal features with very-high-frequency ultrasound [88]. Thus, AMI can be used to investigate the internal defects of a flip chip package, including the following:

- variable underfill flow speed,
- insufficient quantity of fluid underfill,
- BGA (ball-grid array) with underfill,
- PQFP (plastic quad flat pack) with peripheral lead frame delaminations, and
- PQFP with multiple defects.

Ultrasound pulsed into a mounted plastic-encapsulated IC is reflected by internal interfaces, such as the interface between the molding compound and the die face, or the interface between the die attach and the molding compound [89]. Internal defects, e.g., delaminations, voids, and cracks, will cause far stronger reflections and thus will be imaged in higher contrast. Thus, AMI can ensure that the components have been properly mounted and will perform without failure over the long term.

Sonar profile systems are usually employed to explore the elements in the pipeline that are submerged below the water line [90]. These may include submerged debris in the pipe (sewers), grease level (sewers), differential settling, and other submerged deformations and defects. A sonar system consists of an underwater scanner unit, collapsible sonar siphon float, sonar processor/monitor, skid set, and all necessary interconnect cables. It sends a pulse approximately every 1.5 s and each pulse provides an outline of the cross section of the submerged part of the pipe. Accurate measurements can be performed based on these outlines.

Air-coupled ultrasound was employed to generate a tomographic imaging of polymer-based soft drink bottles for the detection of liquid level or foreign objects [91]. Capacitance transducers with polymer membranes are used to generate ultrasonic signals in air over a reasonable bandwidth. And ultrasonic pulse compression (UPC) is then applied to increase the sensitivity of signals transmitted through the containers [91]. The tomographic image can be obtained from the peak signal amplitude and reconstructed by comparing each projection to a set of reference data obtained from a sample containing only water. The defect of 10 mm in diameter can be successfully detected.

1.4.4   Gamma(γ)-Ray and X-Ray Imaging

The use of γ-ray radiographic imaging for cargo inspection is described in [92, 93]. A linear push-broom stereo system is employed to conduct the inspection of cargo/vehicle. A 1D detector array of 256 NaI-PMT probes measure the γ-ray photons passing through the cargo. The γ-ray system or the cargo/vehicle moves in a straight line to obtain a 2D scanning of γ-ray images. A stereo matching algorithm based on a free-form deformable registration approach is developed to derive 3D measurement of objects inside the cargo from the acquired 2D scans [92]. With the help of VTK, the 3D measurements of object inside the cargo container can be visualized.

X-ray inspection systems are adopted by semiconductor industry for the inspection of high-density integrated-chip packages with ball-grid arrays (BGAs). These packages have hidden solder joints and thus optical inspection will not be sufficient [94]. X-ray inspection is well suitable for identifying the shorts, voids, opens, misalignments, and solder integrity of area-array chip packages. For real-time X-ray inspection, an X-ray-sensitive video camera is used to convert X-ray into a video image, which eliminates the cost and delays associated with film processing [94].

In normal X-ray-based inspection, the X-ray is detected by a single type of sensor, which converts the shape, density, and other characteristics into an image [95]. In a dual energy imaging system, high-energy image and low-energy image are captured simultaneously by two types of sensors and the two captured images are combined. Thus, the information about both hard and soft objects can be obtained [95]. Dual energy imaging is being used applications such as security check, foreign object detection from grain, fruit, and meat.

In [96], flaws in aluminum castings were detected by integrating information from multiple views from the sequence of X-ray imaging. Each image in the flaw sequence was first classified as a flaw or non-flaw with support vector machines. Then, the classifiers' responses were cascaded into a vector for further classification as a computation of rate of votes. A better performance was observed in the experiments [96].

Medical computed tomography (CT) is being widely used for medical diagnosis. The backscatter computed tomography (BCT) brings such capabilities to the industrial inspection [97, 98]. Different from conventional radiography, BCT can provide a tomographic or cross-sectional slice view of the object from a single side regardless of its physical size [97]. This BCT system has the following characteristics [97]:

- Single side access
- Material independence
- Accurate thickness gauging
- No material preparation
- Easily interpreted images

Various industrial inspection tasks, such as civil infrastructure, explosive material detection, aerospace and naval applications, and food and agriculture, can benefit from this technique, especially for the applications which are costly to interrupt the services.

1.4.5 Terahertz (THz) Imaging

The spectrum of Terahertz imaging resides between microwave and infrared. With the development of pulsed- and continuous-wave THz systems, it becomes possible to apply THz inspection in various industrial applications, for instance, nondestructive evaluation and process control [99]. The pulsed THz systems make use of electromagnetic picosecond pulses, while the continuous system employs a single-frequency wave at a time [99]. The continuous-wave systems offer a high-frequency resolution in comparison with the pulsed systems as well as phase information for analysis. In contrast, pulsed systems acquire a broad spectral information in a single measurement and thus are suitable to exploring thickness-related information. As THz is strongly absorbed by water, this may limits its applications in determining the hydration state of object under inspection [99]. Moreover, THz wave is not strong enough to penetrate metal layer even of a few microns. Food products have found the varieties of THz for quality inspection. For example, THz imaging can be used to reveal potentially hazardous inclusions in chocolate bar [100]. THz imaging has also been applied to security check and concealed weapon detection [101].

1.4.6 Millimeter-Wave Imaging

The frequency range of millimeter waves (wavelength: 1–10 mm) is from 30 to 300 GHz, which can penetrate nonmetal materials such as clothes, plastics, and mist [102]. Millimeter-wave imaging is applied for the security surveillance and nondestructive inspection of civil infrastructure [103, 104]. Researchers at NTT Laboratories utilized near-field imaging features to detect objects in the sub-millimeter range from the millimeter-wave imaging of concrete structures [102]. For the surface cracks on concrete structures, active imaging, which uses external millimeter-wave radiation, needs to be employed [102]. A crack scanner was developed and successfully applied to the concrete pole covered with bill-posting prevention sheet [105]. A broadband millimeter-wave imaging prototype was described in [106]. Potential applications may expand to the food contamination inspection and security check.

1.4.7 Multispectral Imaging

Multispectral imaging incorporates more than two different types of detectors in one single camera or system. With wafer-level coating of multispectral filters, advanced multispectral cameras offer new opportunities in high-speed line-scan imaging [107]. With a thin-film interference filter technology, a range of wavelengths are selected to pass the multilayer thin films while reflecting others [107]. Multispectral imaging has been applied to automatic optical inspection (AOI), such as security print inspection, electronics manufacturing, food an materials sorting, and medicine [107]. A multispectral RGB + NIR camera was used to inspect both color and NIR images printed with security inks on the banknotes [107]. In electronic circuit manufacturing, multispectral technology is adopted to capture the information at different depths. For example, the inspection of a PCB with three different light sources, i.e., UV, visible, and NIR, can reveal the details of soldering pads and surface morphology, surface, and the wire below the paint to some extent, and copper wires covered by the paint, respectively. Thus, a multispectral camera, which can cover a wide spectrum, is preferred.

In [108], visible and infrared cameras were employed to capture both physical and thermal conditions of railroad equipment undercarriages. The multispectral imaging enhances normal visual inspections for missing or damaged components. In addition, the diagnosis of incipient failures become possible with the integration of the visible and thermal views. In [109], multispectral imaging was used to monitor the seed germination process for the plant quality control. Reference [110] reported the use of a multispectral camera, which is able to acquire visible and near-infrared images from the same scene, to check the quality of citrus. The size, color, and defects in citrus, e.g., oranges, mandarins, and lemons, can be inspected with a higher inspection speed.

1.4.8 Multimodal Inspection System

Multimodal inspection employs multiple inspection methods for a complete characterization of the anomalies or discontinuities [111]. To interpret or analyze the collected inspection data, the data fusion technique need to be applied. In [112], Xiao investigated the use of color imaging, laser range profiling, and X-ray imaging for the detection of critical quality features for wood products. With this multi-imaging system, wood surface features (knots, splits, stains), geometry features (wane, thin board), and internal features (voids, knots) can be detected, which is not possible for an unimodal imaging system.

The inspection of roof tiles for potential anomalies such as fractured edges, cracks, inclusions, and surface defects employed a machine vision system [113, 114]. Images captured by two cameras were stitched before being further processed. A triangulation-based laser scanner was also used in the system to capture the 3D

profile of the tiles. In addition, prior to the vision inspection, an acoustic system recorded the signature of the spectrum initiated by a special hammer and compared it with the baseline reference [115]. Impact acoustic is an effective nondestructive evaluation method for the detection of cooling cracks and internal crack in the tiles [116]. A computational approach based on the principal component analysis for acoustic signal characterization was described in [117].

In the inspection of semiconductor integrated-chip (IC) packages, the solder joints of ball-grid arrays (BGAs) are hidden and thus automated optical inspection (AOI) is not sufficient to ensure solder integrity [118]. Therefore, X-ray inspection is introduced to check area-array and IC components. The combination of AOI and X-ray imaging technologies results in an integrated test system, which is capable of cost-effective inspection of assembled printed circuit boards (PCB). The AOI systems target missing parts, wrong components, polarization errors, part-placement errors, and solder defects. However, AOI systems can only access the visible features. For hidden connection devices, such as BGAs, microBGAs, and flip chips, inspection and analysis with X-ray systems are essential to verify the integrity of hidden connections. An integrated system incorporates both X-ray and vision inspection into one in-line machine at full line speed and offers advantages over each separate technology [118]. The combination system allows to inspect hidden solder joints as well as the PCB assemblies simultaneously for correct parts, position, orientation, and other visual features [119]. A detailed comparison of the AOI- and X-ray-only system with combined AOI/X-ray system is available in [119].

In [35], the combination of high-resolution digital imaging (HRDI) and high-definition video (HDV) digital crack detection and IR thermography technology was implemented for in-service bridge inspections. Thermography is used for locating possible delamination and spalling of concrete.

1.5 Expansion in Applications

Machine vision systems have found numerous applications in various manufacturing processes and food production industry. Table 1.6 summarizes some typical applications in different categories. The list is far from complete. The purpose of this section is to highlight the potential expanding in applications.

To tackle the challenge of inspecting specular free-form surfaces, a vision system consisting of one-line laser and two cameras are configured in a reciprocal deflectometric setup, which is applicable to typical surface defects on moving glazed ceramic tiles [120]. The reflection of the line laser is captured on a translucent screen. Surface defects are presented as gaps or bulges. This system is able to detect, characterize, and localize a range of typical surface defects on specular surfaces. In [121], the authors proposed a procedure to detect the detects from strongly reflected metal surface, where the acquired image is represented in wavelet transform domain and spectral measures are used to classify the defect with a

Table 1.6 Typical machine vision applications

Industry	Applications	References
Food/pharmaceutical	Inspection of the internal and external quality of fruits and vegetables	[81, 132, 140]
	Food storage container assembly and inspection	[81, 91, 110, 141]
	Detect if bags of cheddar cheese have been vacuum sealed properly	[142]
Automotive	Check car fenders, inspect canister	[143, 144]
	Vehicle body spot welding inspection	[145]
	Evaluation and visualization of surface defects on autobody panels	[146]
	Inspection of vacuum florescent displaying	[147]
	Drive shafts, brake pads, color calipers, arm nuts	[148]
Aerospace	Inspection of aircraft composite materials	[149]
	Surface (pillowing) deformation due to hidden corrosion	[58]
Railway	Joint bar inspection	[150, 151]
	Railway track components	[152, 153]
	Inspection of missing fastening components in high-speed railway	[154]
	Wayside inspection for the mechanical components of railcar	[155, 156]
Semiconductor	Diagnosis of internal defects of flip chip package	[88]
Electronic component	Defect detection of printed circuit board assemblies	[119]
	Inspection of printed circuit board	[157]
Plastic/Rubber/Paper	Detect deviations of rapidly moving paper web	[158]
Forestry	Lumber defect detection and inspection for quality control	[112, 159]

support vector machine. A morphological processing was proposed in [122] to detect structural defects on bumpy metallic surface, where genetic algorithms were employed to automatically learn morphology processing parameters such as structuring elements and defect segmentation threshold. An automated defect inspection system for highly specular reflection curved surfaces, such as chrome-plated surfaces, was introduced in [123]. In this vision system, a set of images are captured under different illumination with an optimized relative position and illumination parameter. A synthetic image with specular spots removed is obtained from this set of images. With the template matching and morphological processing, defects can be extracted from the composite image. More detailed description about this system can be found in another chapter of this book.

The generation and use of 3D images for inspection and quality control becomes increasingly demanded in mechanical and manufacturing industry [124]. A review of 3D sensing techniques and devices as well as various applications is represented

in [124]. The major strengths and weaknesses of optical range imaging as listed below [124] were discussed.

- Laser triangulators,
- Structured light,
- Stereo vision,
- Photogrammetry,
- Time of flight,
- Interferometry
- Moiré fringe range contours,
- Shape from focusing,
- Shape from shadows,
- Texture gradients,
- Shape from shading, and
- Shape from photometry.

3D imaging covers a wide range of applications. The condition assessment of different modes of transportation, including highways and roads, airport runways, tunnels and rail, was conducted with 3D laser-based technology [125]. Usually, a regular tunnel inspection needs to be conducted for detecting abnormalities such as cracks, scales, spalls, and corrosion at early stage. Traditional method with manual recording from mobile scaffolding is labor-intensive and time-consuming [126]. A 3D laser system, named "Tunnelings," was developed, which consists of multiple 3D laser profilers, high-speed industrial cameras, and sophisticated image processing tools [127]. For the automated inspection of tunnels, 2D images, and high-resolution 3D profiles of infrastructure surfaces are collected and can serve as a permanent record. The accurate 3D information allows the tunnel lining inspection and structural evaluation [128]. A similar system named "tCrack" was developed to assure the safety of Swiss Gotthard Base Tunnel [126]. The "tCrack" system consists of ten cameras mounted on a site vehicle with or without tracks. While the vehicle is driven along the center of the tunnel, the ten cameras scan the entire walls and surfaces. The strips captured by each camera are then merged into continuous overlapping images to create a complete picture of the tunnel for further analysis and assessment. 3D laser profiling has also been applied to assess joint concrete pavements [129].

Stereo cameras were used to visualize the deformation of aircraft propellers [130]. Two CMOS cameras were arranged in stereoscopic manner to capture image pairs for 3D surface shape. A random dot pattern as well as checker board markers is applied to the surface of propeller to provide key reference points. By matching those key points from two images, the shape of surface can be obtained. The in-flight images need to be de-warped first and cross-related for image pattern correlation, where the 3D coordinates of the same dot pattern can be determined through a triangulation process.

Grading and estimating the quality of fruits and vegetables employs machine vision systems covering both visible and non-visible spectrum, which makes it possible to identify the presence of defects through external and internal features

from acquired images. An excellent review of machine vision technologies for fruits and vegetables is available in [131, 132].

1.6 Summary

This chapter reviews the major advances of machine vision technology for industrial inspection. The significant progresses in solid-state imaging sensors, LED illumination devices, high-performance processors, networking and communication technologies, human–computer interfaces, and artificial intelligence enrich the functionalities of each component in the machine vision system and offer great flexibility in the overall system configuration. To tackle the inspection challenges in future manufacturing, the adoption of new smart cameras and new imaging devices as well as the development of new algorithms for image understanding and interpretation become paramount and essential. Combined technologies and multimodal inspection need the fusion of heterogeneous information from different sources, even though the input is not visual information. Thus, a common data platform or data cloud needs be developed or established to align and manipulate huge volumes of different types of inspection data. A more intelligent vision in industrial inspection is foreseen.

References

1. Dowling KJ, Mueller GG, Lys IA (2003) Systems and methods for providing illumination in machine vision systems. Patent: US 6624597B2. United States, Boston, MA, US: Assignee: Color Kinetics, Inc
2. Newman TS, Jain AK (1995) A survey of automated visual inspection. Comput Vis Image Underst 61(2):231–262. doi:10.1006/cviu.1995.1017
3. Batchelor BG (1999) Coming to terms with machine vision and computer vision. Advanced imaging, pp. 22–26
4. Labudzki R, Legutko S (2011) Applications of machine vision. Manufact Ind Eng 2:27–29
5. Malamas EN, Petrakis EG, Zervakis M, Petit L, Legat JD (2003) A survey on industrial vision systems, applications and tools. Image Vis Comput 21(2):171–188. doi:10.1016/S0262-8856(02)00152-X
6. Thomas A, Rodd M, Holt J, Neill C (1995) Real-time industrial visual inspection: a review. Real-Time Imaging 1(2):139–158. doi:10.1006/rtim.1995.1014
7. Gregory R Planning a pc-based machine vision system. Online white paper. URL http://archives.sensorsmag.com/articles/0498/vis0498/index.htm. Accessed in March 2015
8. Bier J (2011) Implementing vision capabilities in embedded systems. Online white paper (2011). URL http://www.embedded-vision.com. Accessed in March 2015
9. Gardner JS Challenges to embedding computer vision. Online. URL http://www.embedded-vision.com. Accessed in March 2015
10. Berkeley Design Technology Inc (2011) Implementing vision capabilities in embedded systems. Presentation at 2011 Embedded Systems Conference Silicon Valley
11. Dipert B, Khan K (2013) The embedded vision revolution. Online article (2013). URL http://www.mddionline.com/article/embedded-vision-revolution. Accessed in March 2015

12. Wolf W, Ozer B, Lv T (2002) Smart cameras as embedded systems. Computer 35(9):48–53. doi:10.1109/MC.2002.1033027

13. Shi D, Lichman S (2005) Smart cameras: a review. In: Proceedings of 2005 Asia-Pacific workshop on visual information processing. Hong Kong, China, pp 11–13

14. de Sousa A (2003) Smart cameras as embedded systems. In: Proceedings first international conference on computer applications, pp. 105–112. Session 4: Embedded Systems

15. Shi Y, Real F (2010) Smart cameras: fundamentals and classification. In: Belbachir AN (ed) Smart cameras. Springer, US, pp 19–34. doi:10.1007/978-1-4419-0953-4-2

16. Wilson A (2013) Time-of-flight camera captures VGA images at high speed. Vis Syst Des 18 (1):11–12

17. Malik AW, Thoornberg B, Kumar P (2013) Comparison of three smart camera architectures for real-time machine vision system. Int J Adv Rob Syst 10:1–12. doi:10.5772/57135 10:402

18. Elouardi A, Bouaziz S, Dupret A, Lacassagne L, Klein J, Reynaud R (2006) A smart sensor for image processing: towards a system on chip. In: IEEE International Symposium on Industrial Electronics, vol 4, pp. 2857–2862. doi 10.1109/ISIE.2006.296069

19. Rodriguez-Vazquez A, Dominguez-Castro R, Jimenez-Garrido F, Morillas S (2010) A CMOS vision system on-chip with multicore sensory processing architecture for image analysis above 1000 f/s. In: Bodegom E, Nguyen V (eds) Proceedings of SPIE, sensors, cameras, and systems for industrial/scientific applications XI, vol 7536, pp. 75,3600–75,3600–11. San Jose, California, USA. doi:10.1117/12.839183

20. Martin D (2007) A practical guide to machine vision lighting. Online white paper. URL http://www.graftek.com/. Accessed in March 2015

21. Hecht K (2005) Integrating LED illumination system for machine vision systems. Patent: US6871993B2, Hatfield, PA, US. Assignee: Accu-Sort Systems, Inc. 2006

22. Gardasoft The practical use of LED light controllers within machine vision systems. Online white paper. URL http://www.gardasoft.com. Accessed in March 2015

23. Wang W, Li W (2009) Design of reconfigurable LED illumination control system based on fpga. In: Image and Signal Processing, CISP '09. 2nd International Congress on, pp. 1–4. doi:10.1109/CISP.2009.5304361

24. MICROSCAN Eight tips for optimal machine vision lighting. Online technology white paper. URL http://www.microscan.com. Accessed in March 2015

25. Dowling KJ, Mueller GG, Lys IA (2006) Systems and methods for providing illumination in machine vision systems. Patent: US7042172B2, United States, Boston, MA, US. Assignee: Color Kinetics Inc

26. Dechow D (2013) Explore the fundamentals of machine vision: part I. Vision Sys Des 18 (2):14–15

27. Teledyne Dalsa: Application notes and technology primer: CCD versus CMOS. Online article. URL https://www.teledynedalsa.com. Accessed in March 2015

28. Apler G (2011) CCD versus CMOS image sensors: the lines are blurring. Online article (2011). URL http://info.adimec.com/. Accessed in March 2015

29. Bosiers JT, Peters IM, Draijer C, Theuwissen A (2006) Technical challenges and recent progress in (CCD) imagers. Nucl Instrum Methods Phys Res Sect. A Accelerators, Spectrometers, Detectors and Associated Equipment 565(1):148–156. doi http://dx.doi.org/10.1016/j.nima.2006.05.033. Proceedings of the International Workshop on Semiconductor Pixel Detectors for Particles and Imaging (PIXEL) 2005 International Workshop on Semiconductor Pixel Detectors for Particles and Imaging

30. Teledyna Dalsa: X-ray imaging: emerging digital technology—CMOS detectors. Online white paper. URL https://www.teledynedalsa.com/imaging/products/x-ray/. Accessed in March 2015

31. Korthout L, Verbugt D, Timpert J, Mierop A, de Haan W, Maes W, de Meulmeester J, Muhammad W, Dillen B, Stoldt H et al (2009) A wafer-scale CMOS APS imager for medical X-ray applications. In: International Image Sensor Workshop. Bergen, Norway

32. Princeton Instruments Imaging Group: Introduction to scientific InGaAs FPA cameras. Oneline technical note (2012). URL http://www.princetoninstruments.com/. Accessed in March 2013

33. Barton JB, Cannata RF, Petronio SM (2002) InGaAs NIR focal plane arrays for imaging and DWDM applications. In: AeroSense 2002, International society for optics and photonics, pp. 37–47

34. Hamamatsu Photonics High sensitivity cameras: principle and technology. Online technical note. URL http://www.hamamatsu.com. Accessed in March 2015

35. Matsumoto M, Mitani K, Sugimoto M, Hashimoto K, Miller R (2012) Innovative bridge assessment methods using image processing and infrared thermography technology. IABSE congress report 18(13):1181–1188. URL http://www.ingentaconnect.com/content/iabse/congr/2012/00000018/00000013/art00004

36. Munro JF (2008) Systems for capturing three-dimensional images and methods thereof. Patent: US20080050013A1, United States, Rochester, NY, USA

37. von Fintel R Comparison of the most common digital interface technologies in vision technology: camera link, USB3 vision, GigE vision, FireWire. Online white paper. URL http://www.baslerweb.com. Accessed in March 2015

38. Edmund Optics Imaging electronics 101: camera types and interfaces for machine vision applications. Online white paper. URL http://www.edmundoptics.com/technical-resources-center/imaging/camera-types-and-interfaces-for-machine-vision-applications. Accessed in March 2015

39. Automated Imaging Association Gige vision—true plug and play connectivity. Online article. URL http://www.visiononline.org. Accessed in March 2015

40. Automated Imaging Association Camera link the only real-time machine vision protocol. Online article. URL http://www.visiononline.org. Accessed in March 2015

41. Association U.I.V CoaXPress high speed camera interface. Online white paper. URL www.ukiva.org. Accessed in March 2015

42. Adimec (2013) Multi-camera vision system with CoaXPress. Online white paper (2013). Accessed in March 2015

43. Matrox GPU processing using MIL. Online white paper. URL http://www.matrox.com. Accessed in March 2015

44. Tweddle B (2009) Graphics processing unit (GPU) acceleration of machine vision software for space flight applications. Workshop on space flight software. URL http://flightsoftware.jhuapl.edu. Presentation Slide

45. MVTec (2011) MVTec improves factory throughput and quality using NVIDIA GPU-accelerated inspection automation. Online Article. URL http://www.mvtec.com. Accessed in March 2015

46. Larson B (2015) GPU-accelerated machine vision. Camera and photos 21. URL http://www.objc.io/issue-21/gpu-accelerated-machine-vision.html

47. Dipert B, Alvarez J, Touriguian M (2012) Embedded vision: FPGAs' next notable technology opportunity. Xcell J 78:14–19

48. Chen YC, Wang YT (2008) Development of a low-cost machine vision system and its application on inspection processes. Tamkang J Sci Eng 11(4):425–431

49. Besiris D, Tsagaris V, Fragoulis N, Theoharatos C (2012) An FPGA-based hardware implementation of configurable pixel-level color image fusion. IEEE Trans Geosci Remote Sens 50(2):362–373. doi:10.1109/TGRS.2011.2163723

50. Baumgartner D, Roessler P, Kubinger W, Zinner C, Ambrosch K (2009) Benchmarks of low-level vision algorithms for DSP, FPGA, and mobile PC processors. In: Kisacanin B, Bhattacharyya S, Chai S (eds) Embedded computer vision, advances in pattern recognition. Springer, London, pp. 101–120. doi:10.1007/978-1-84800-304-0-5

51. Dechow D (2013) Explore the fundamentals of machine vision: part II. Vision Sys Des 18(4):16–18

52. Hornberg A (ed) (2006) Handbook of machine vision. 978-3-527-40584-8. Wiley-VCH, Favoritenstrasse 9/4th Floor/1863

53. Batchelor BG, Whelan PF (1997) Intelligent vision systems for industry. 3540199691. Springer, London
54. Jain R, Kasturi R, Schunck BG (1995) Machine vision. No. 0-07-032018-7 in McGraw-Hill Series in computer science. McGraw-Hill, Inc., New York, USA
55. Pernkopf F (2005) 3D surface acquisition and reconstruction for inspection of raw steel products. Comput Ind 56(89):876–885. doi:http://dx.doi.org/10.1016/j.compind.2005.05.025. (Machine Vision Special Issue)
56. Pernkopf F (2004) Detection of surface defects on raw steel blocks using bayesian network classifiers. Pattern Anal Appl 7(3):333–342. doi:10.1007/BF02683998
57. Liu Z, Genest M, Marincak A, Forsyth D (2008) Characterization of surface deformation with the edge of lighttm technique. Mach Vis Appl 19(1):35–42. doi:10.1007/s00138-007-0075-1
58. Liu Z, Genest M, Forsyth D, Marincak A (2009) Quantifying surface deformation with the Edge of Light enhanced visual inspection. Instrum Measur IEEE Transactions on 58(2):416–422. doi:10.1109/TIM.2008.2003312
59. Lillywhite K, Lee DJ, Tippetts B, Archibald J (2013) A feature construction method for general object recognition. Pattern Recogn 46(12):3300–3314. doi:10.1016/j.patcog.2013.06.002
60. Motoda H, Liu H (2002) Feature selection, extraction and construction. Commun Inst Inf Comput Mach 5:67–72
61. Wikipedia: Convolutional neural network. URL http://en.wikipedia.org/wiki/Convolutional_neural_network. Accessed in March 2015
62. Theano Development Team: Deeplearning 0.1 document: Convolutional neural networks (lenet). Online document. URL http://deeplearning.net/tutorial/lenet.html. Accessed in March 2015
63. Ciresan D, Meier U, Schmidhuber J (2012) Multi-column deep neural networks for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3642–3649
64. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: NIPS Proceedings of advances in neural information processing systems, pp. 1097–1105
65. OpenCV: Open source computer vision. Website: URL http://opencv.org/
66. SimpleCV (2013) Computer vision platform using Python. Website: URL http://simplecv.org/. Accessed in March 2015
67. Kitware Inc. (2014) Visualization toolkit. URL http://www.vtk.org/
68. Gould S (2012) Darwin: A framework for machine learning and computer vision research and development. J Mach Learn Res 13:3533–3537
69. Grigorescu SM, Ristic-Durrant D, Graser A. (2009) ROVIS: Robust machine vision for service robotic system FRIEND. In: Intelligent robots and systems IROS 2009. IEEE/RSJ International Conference on pp. 3574–3581.doi:10.1109/IROS.2009.5354596
70. Kaehler A, Bradski G (2013) Learning OpenCV: computer vision in C++ with the OpenCV Library, 1st edn. 978-0-596-51613-0. O'Reilly Media, Sebastopol, CA, USA
71. Demaagd K, Oliver A, Oostendorp N, Scott K (2012) Practical computer vision with SimpleCV, 1st edn. 978-1-449-32036-2. O'Reilly Media, Sebastopol, CA, USA
72. Schroeder WJ, Martin K, Lorensen W (2003) The visualization toolkit: an object-oriented approach to 3D graphics, 3rd edn. Kitware, Inc. (formerly Prentice-Hall), USA
73. Schroeder WJ, Martin K, Lorensen WE (1996) The design and implementation of an object-oriented toolkit for 3D graphics and visualization. In: Proceedings of the 7th conference on visualization '96, VIS '96. IEEE Computer Society Press, Los Alamitos, CA, USA, pp 93–111. URL http://dl.acm.org/citation.cfm?id=244979.245018
74. Eledath J (2013) Tools for democratizing computer vision: automated performance characterization. Embedded Vision Summit East Presentations. Westford, MA, USA
75. Solid State Division Characteristics and use of infrared detectors. Online. URL http://www.hamamatsu.com. Retrieved in March 2015

76. Maldague XPV (2001) Theory and practice of infrared technology for nondestructive testing. wiley series in microwave and optical engineering. Wiley, USA

77. Rybicki GB, Lightman AP (1979) Radiative Processes in Astrophysics. Wiley, New York

78. Blum RS, Liu Z (eds) (2005) Multi-sensor image fusion and its applications. signal processing and communications. Taylor and Francis, UK

79. Rogalski A (2002) Infrared detectors: an overview. Infrared Phys Technol 43(35):187–210. doi:10.1016/S1350-4495(02)00140-8

80. Tech note (2015) IR lighting (NIR—near infrared). Online (2015). URL http://smartvisionlights.com

81. Vadivambal R, Jayas D (2011) Applications of thermal imaging in agriculture and food industrya review. Food Bioprocess Technol 4(2):186–199. doi:10.1007/s11947-010-0333-5

82. Bagavathiappan S, Lahiri B, Saravanan T, Philip J, Jayakumar T (2013) Infrared thermography for condition monitoring a review. Infrared Phys Technol 60:35–55. doi:10.1016/j.infrared.2013.03.006

83. Pinter M (2015) Advance in UV light for machine vision applications. Online (2015). URL http://smartvisionlights.com

84. Richards A (2006) UV imaging opens new applications. Vision Systems Design

85. Wilson A (2012) Enhanced cameras detect targets in the UV spectrum. Vis Syst Des 17 (10):13–14

86. Slaughter D, Obenland D, Thompson J, Arpaia M, Margosan D (2008) Non-destructive freeze damage detection in oranges using machine vision and ultraviolet fluorescence. Postharvest Biol Technol 48(3):341–346. doi:10.1016/j.postharvbio.2007.09.012

87. Kondo N, Kuramoto M, Shimizu H, Ogawa Y, Kurita M, Nishizu T, Chong VK, Yamamoto K (2009) Identification of fluorescent substance in mandarin orange skin for machine vision system to detect rotten citrus fruits. Eng Agric Environ Food 2(2):54–59. doi:10.1016/S1881-8366(09)80016-5

88. Adams T (2000) What happened here: diagnosis of internal defects from acoustic images. Online white paper (2000). URL http://www.satech8.com. Accessed in March 2015

89. Adams T (2002) Acoustic micro imaging finds hidden defects. Online article (2002). URL http://www.sonoscan.com. Accessed in March 2015

90. Liu Z, Kleiner Y, Rajjani B, Wang L, Condit W (2012) Condition assessment of water transmission and distribution systems. Tech. Rep. EPA/600/R-12/017, United States Environmental Protection Agency, National Risk Management Research Laboratory, Ottawa, Ontario, Canada. (Institute for Research in Construction, National Research Council Canada)

91. Gan T, Hutchins D, Billson D (2002) Preliminary studies of a novel air-coupled ultrasonic inspection system for food containers. J Food Eng 53(4):315–323. doi:10.1016/S0260-8774(01)00172-8

92. Zhu Z, Hu YC, Zhao L (2010) Gamma/x-ray linear pushbroom stereo for 3D cargo inspection. Mach Vis Appl 21(4):413–425. doi:10.1007/s00138-008-0173-8

93. Zhu Z, Zhao L, Lei J (2005) 3D measurements in cargo inspection with a gamma-ray linear pushbroom stereo system. In: Computer Vision and Pattern Recognition—Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on, pp 126–126. doi:10.1109/CVPR.2005.380

94. Amtower R (2000) X-ray enhanced AOI. Online white paper. URL http://www.satech8.com/. Accessed in March 2015

95. OPTO-Semiconductor HandbooK, chap. X-ray detectors, pp. 1–21. Hamamatsu. Accessed in March 2015

96. Pieringer C, Mery D (2010) Flaw detection in aluminum die castings using simultaneous combination of multiple views. Insight 52(10):548–552

97. Backscatter computed tomography (2015) Website: URL http://www.inversasystems.com. Accessed in March 2015

98. Khan M, Liu Z (2013) Backscatter computed tomography technique for the detection of damage in pipes. Technical report LTR-SMPL-2013-0071, National Research Council Canada. Ottawa, Ontario, Canada

99. Jansen C, Scherger B, Jordens C, Al-Naib IAI, Koch M Terahertz imaging spectroscopy for quality inspection in the food industry. Online article. URL http://www.labint-online.com. Accessed in March 2015

100. Jördens C, Rutz F, Koch M (2006) Quality assurance of chocolate products with terahertz imaging. In: European Conference on NDT. Berlin, German, pp. 25–29

101. Blasch E, Liu Z, Petkie D, Ewing R, Pomrenke G, Reinhardt K (2012) Image fusion of the terahertz-visual naecon grand challenge data. In: Aerospace and Electronics Conference (NAECON), 2012 IEEE National, pp. 220–227. doi 10.1109/NAECON.2012.6531058

102. Oka S, Mochizuki S, Togo H, Kukutsu N (2009) Inspection of concrete structures using millimeter-wave imaging technology. NTT Tech Rev 7(3):1–6

103. Oka S, Togo H, Kukutsu N, Nagatsuma T (2008) Latest trends in millimeter-wave imaging technology. Prog Electromagnet Res Lett 1:197–204

104. Sheen DM, McMakin DL, Collins HD, Hall TE, Severtsen RH (1996) Concealed explosive detection on personnel using a wideband holographic millimeter-wave imaging system. In: Kadar I, Libby V (eds) Proceedings of SPIE: signal processing, sensor fusion, and target recognition, vol 2755. Orlando, FL, USA, pp. 503–513. doi:10.1117/12.243191. URL http://dx.doi.org/10.1117/12.243191

105. adn Takafumi Kojima HT, Mochizuki S, Kukutsu N (2012) Millimeter-wave imaging for detecting surface cracks on concrete pole covered with bill-posting prevention sheet. NTT Technical Review 10(2):1–6

106. Mizuno M (2008) Broadband millimeter wave imaging system. J Natl Inst Inf Commun Technol 55(1):53–59

107. He X (2015) Multispectral imaging extends vision technology capability. Photonics Spectra, USA, pp 1–4

108. Hart J, Resendiz E, Freid B, Sawadisavi S, Barkan C, Ahuja N (2008) Machine vision using multi-spectral imaging for undercarriage inspection of railroad equipment. In: Proceedings of the 8th world congress on railway research. Seoul, Korea

109. Falkenstein P (2012) Multispectral imaging plants roots in quality control. Vision Sys Des 17(11):23–25

110. Aleixos N, Blasco J, Navarrn F, Molt E (2002) Multispectral inspection of citrus in real-time using machine vision and digital signal processors. Comput Electron Agric 33(2):121–137. doi:10.1016/S0168-1699(02)00002-9

111. Liu Z, Forsyth DS, Komorroski JP, Hanasaki K, Kiruba K (2007) Survey: State of the art of NDE data fusion. IEEE Trans Instrum Meas 56(6):2435–2451

112. Xiao X (1998) A multiple sensors approach to wood defect detection. Doctor dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

113. ibea Hamburg GmbH: Surface anomaly inspection system. Online white paper. URL http://www.ibea.de/. Accessed in March 2015

114. Wilson A (2014) Roof tiles inspected with sound and vision. Vis Syst Des 19(3):16–17

115. ibea Hamburg GmbH Acoustic crack inspection for clay tiles. Online white paper. URL http://www.ibea.de/. Accessed in March 2015

116. Luk B, Liu K, Jiang Z, Tong F (2009) Robotic impact-acoustics system for tile-wall bonding integrity inspection. Mechatronics 19(8):1251–1260. doi:http://dx.doi.org/10.1016/j.mechatronics.2009.07.006. URL http://www.sciencedirect.com/science/article/pii/S0957415809001329

117. Tong F, Tso S, Hung M (2006) Impact-acoustics-based health monitoring of tile-wall bonding integrity using principal component analysis. J Sound Vib 294(12):329–340. doi:10.1016/j.jsv.2005.11.017

118. Amtower R X-ray enhanced AOI. Online white paper. URL http://www.satech8.com. Accessed in March 2015

119. LeBlond C Combining AOI and AXI: the best of both worlds. Online white paper. URL http://www.satech8.com. Accessed in March 2015

120. Wedowski RD, Atkinson GA, Smith ML, Smith LN (2012) A system for the dynamic industrial inspection of specular freeform surfaces. Opt Lasers Eng 50(5):632–644. doi:10.1016/j.optlaseng.2011.11.007

121. Zhang X, Ding Y, Lv Y, Shi A, Liang R (2011) A vision inspection system for the surface defects of strongly reflected metal based on multi-class SVM. Expert Syst Appl 38(5):5930–5939. doi:10.1016/j.eswa.2010.11.030

122. Zheng H, Kong L, Nahavandi S (2002) Automatic inspection of metallic surface defects using genetic algorithms. J Mater Process Technol 125126:427–433. doi:10.1016/S0924-0136(02)00294-7

123. Li CJ, Zhang Z, Nakamura I, Imamura T, Miyake T, Fujiwara H (2012) Developing a new automatic vision defect inspection system for curved ssurface with hihigh specular reflection. Int J Innovative Comput Inf Control 8(7):5121–5136

124. Sansoni G, Trebeschi M, Docchio F (2009) State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation. Sensors 9(1):568–601. doi:10.3390/s90100568

125. Habel R, Laurent J, Hebert JF, Talbot M, Fox-Ivey R (2013) Use of 3D scanning technology for automated inspection of multi-modal transportation infrastructure. In: 17th IRF world meeting and exhibition. Riyadh, Saudi Arabia, pp. 1–20

126. Basler Making tunnels safer—basler pilot gige cameras reliably detect cracks in tunnels. Online white paper. URL http://www.baslerweb.com/. Accessed in March 2015

127. Gavilan M, Sanchez F, Ramos JA, Marcos O (2013) Mobile inspection system for high-resolution assessment of tunnels. In: The 6th international conference on structural health monitoring of intelligent infrastructure. Hong Kong, China, pp. 1–10

128. Laurent J, Fox-Ivey R, Dominguez FS, Garcia JAR (2014) Use of 3D scanning technology for automated inspection of tunnels. In: Proceedings of the world tunnel congress 2014. Foz do Iguau, Brazil

129. Tsai Y, Li F (2012) Critical assessment of detecting asphalt pavement cracks under different lighting and low intensity contrast conditions using emerging 3d laser technology. J Trans Eng 138(5):649–656. doi:10.1061/(ASCE)TE.1943-5436.0000353

130. Boden F, Stasicki B (2015) Stereo camera visualize propeller. Vis Sys Des 20(2):21–25

131. Brosnan T, Sun DW (2004) Improving quality inspection of food products by computer visiona review. J Food Eng 61(1):3–16. doi:10.1016/S0260-8774(03)00183-3 Applications of computer vision in the food industry

132. Cubero S, Aleixos N, Molt E, Gmez-Sanchis J, Blasco J (2011) Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. Food Bioprocess Technol 4(4):487–504. doi:10.1007/s11947-010-0411-8

133. Fermum L (2015) Solutions for industrial machine vision. Online article. URL http://www.vision-doctor.co.uk. Accessed in March 2015

134. Matrox Vision library or vision-specific IDE (2012) which is right for you? matrox imaging white paper. Retrieved in March 2015

135. National Instruments, Austin, Texas USA (2009) NI Vision: NI Vision Builder for Automated Inspection Tutorial. Accessed in March 2015

136. MVTec Software GmbH, Munchen, Germany: HALCON the power of machine vision: quick guide

137. MVTec Software GmbH: MERLIC—a new generation of vision software. Website. URL http://www.mvtec.com/products/merlic/. Accessed in March 2015

138. Tordivel, Oslo, Norway: scorpion vision software version X product overview. Product brochure

139. Teledyne Dalsa, Boston, USA: Sapera Vision Software: Sapera Essential. Accessed in March 2015

140. Ruiz-Altisent M, Ruiz-Garcia L, Moreda G, Lu R, Hernandez-Sanchez N, Correa E, Diezma B, Nicola B, Garc-a-Ramos J (2010) Sensors for product characterization and quality

of specialty cropsa review. Comput Electron Agric 74(2):176–194. doi:10.1016/j.compag.2010.07.002

141. Mahalik NP, Nambiar AN (2010) Trends in food packaging and manufacturing systems and technology. Trends Food Sci Technol 21(3):117–128. doi:10.1016/j.tifs.2009.12.006 Advances in Food Processing and Packaging Automation

142. Flood N, Bailey B (2013) Vision helps dairy spot slack cheese bags. Vision Sys Des 18 (9):17–22

143. Walker C (2014) Filters reduce glare in automotive canister inspection. Vis Syst Des 19 (5):17–19

144. Wilson A (2015) Robotic vision system checks car fenders. Vis Syst Des 20(1):9–11

145. Zhou J, Lee I, Thomas B, Menassa R, Farrant A, Sansome A (2011) Applying spatial augmented reality to facilitate in-situ support for automotive spot welding inspection. In: Proceedings of the 10th international conference on virtual reality continuum and its applications in industry, VRCAI '11. ACM, New York, NY, USA, pp 195–200. doi:10.1145/2087756.2087784

146. Andersson A (2009) Evaluation and visualisation of surface defects on auto-body panels. J Mater Process Technol 209(2):821–837. doi:10.1016/j.jmatprotec.2008.02.078

147. Lu Y, Tie-Qi YL, Chen J, Tisler A, Zhang J (2000) An advanced machine vision system for VFD inspection. In: PCM 2000, pp. 1–6

148. Yardley E (2015) Vision system inspects automotive sub-assemblies. Vis Syst Des 20(2):14–19

149. Duan Y, Servais P, Genest M, Ibarra-Castanedo C, Maldague X (2012) ThermoPoD: a reliability study on active infrared thermography for the inspection of composite materials. J Mech Sci Technol 26(7):1985–1991. doi:10.1007/s12206-012-0510-8

150. Berry A, Nejikovsky B, Gibert X, Tajaddini A (2008) High speed video inspection of joint bars using advanced image collection and processing techniques. In: Proceedings of world congress on railway research. Seoul, Korea, pp. 1–13

151. Gibert-Serra X, Berry A, Diaz C, Jordan W, Nejikovsky B, Tajaddini A (2007) A machine vision system for automated joint bar inspection from a moving rail vehicle. In: ASME/IEEE joint rail conference and internal combustion engine spring technical conference. Pueblo, Colorado, USA

152. Resendiz E, Hart J, Ahuja N (2013) Automated visual inspection of railroad tracks. Intell Transp Syst IEEE Trans on 14(2):751–760. doi:10.1109/TITS.2012.2236555

153. Resendiz E, Molina L, Hart J, Edwards R, Sawadisavi S, Ahuja N, Barkan C (2010) Development of a machine vision system for inspection of railway track components. In: 12th world conference on transport research. Lisbon, Portugal, pp. 1–22

154. Zhang H, Yang J, Tao W, Zhao H (2011) Vision method of inspecting missing fastening components in high-speed railway. Appl Opt 50(20):3658–3665. doi:10.1364/AO.50.003658

155. Hart JM, Ahuja N, Barkan C, Davis DD (2004) A machine vision system for monitoring railcar health: Preliminary results. Technology Digest (TD-04-008), pp 1–4

156. Schlake BW, Todorovic S, Edwards JR, Hart JM, Ahuja N, Barkan CP (2010) Machine vision condition monitoring of heavy-axle load railcar structural underframe components. Proc Inst Mech Eng Part F J Rail Rapid Transit 224(5):499–511

157. Chen TQ, Zhang J, Zhou Y, Murphey YL (2001) A smart machine vision system for PCB inspection. In: Proceedings of engineering of intelligent systems, 14th international conference on industrial and engineering applications of artificial intelligence and expert systems, IEA/AIE. Budapest, Hungary, pp. 513–518. doi 10.1007/3-540-45517-5-57

158. Ruuska H (2009) Method for monitoring a rapidly-moving paper web and corresponding system

159. Qiu Z (1996) A simple machine vision system for improving the edging and trimming operations performed in hardwood sawmills. Master thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

Part I
Technology Advances

Chapter 2
Infrared Vision: Visual Inspection Beyond the Visible Spectrum

Clemente Ibarra-Castanedo, Stefano Sfarra, Marc Genest and Xavier Maldague

Abstract Infrared (IR) vision for the nondestructive testing (NDT) of materials has considerably grown in the last few years as a results of the continual technological progress and the development of advanced signal processing techniques. Typical applications can be divided into two groups: (1) reflectography/transmittography in the near (NIR) and short-wave (SWIR) infrared spectra, which have been recently applied in the inspection of semitransparent composite materials such as glass, aramid–phenolic, and a variety of natural fibers; and (2) infrared thermography, which involves the detection of surface and subsurface defects based on the differences in thermal signatures in the mid-wave (MWIR) and long-wave (LWIR) infrared bands. Infrared reflectography and thermography have shown interesting complementarities between them as well as with other NDT methods such as ultrasound testing, eddy current testing, and holographic interferometry. In this chapter, a review of classical and recent applications of infrared vision is provided and discussed in detail with examples.

C. Ibarra-Castanedo (✉) · X. Maldague
Computer Vision and Systems Laboratory, Université Laval, 1065, av. de la Médecine, Quebec G1V 0A6, Canada
e-mail: IbarraC@gel.ulaval.ca

X. Maldague
e-mail: MaldagX@gel.ulaval.ca

S. Sfarra
Las.E.R. Laboratory, Department of Industrial and Information Engineering and Economics (DIIIE), University of L'Aquila, Piazzale E. Pontieri no. 1, 67100 Monteluco di Roio—L'Aquila, AQ, Italy
e-mail: Stefano.Sfarra@univaq.it

M. Genest
National Research Council Canada, 1200 Montreal Road, Bldg. M-14, Room 130, Ottawa, ON K1A 0R6, Canada
e-mail: Marc.Genest@nrc-cnrc.gc.ca

Contents

2.1   Infrared Vision

Infrared (IR) vision can be defined as a subdivision of computer vision, i.e., the use of computers to emulate human vision [1], which employs low-, mid-, and high-level computerized processes to analyze images generated in the infrared part of the electromagnetic (EM) spectrum. To "analyze images" is a very wide concept, which depends on the user's goal (noise reduction, contrast enhancement, segmentation, etc.) as well as on the way images are generated (a single image, a dynamic sequence, etc.).

The EM spectrum, as depicted in Fig. 2.1, covers all kind of EM radiation, from γ-rays in the shortest wavelengths (in the order of femtometers, e.g., the size of the diameter of a gold nucleus ~ 8.45 fm) end of the spectrum (Fig. 2.1, left), to radio frequencies in the longest (in the order of megameters, e.g, the size of the Earth's equatorial diameter ~ 12.7 Mm) wavelengths (Fig. 2.1, right).

The IR band is located between the visible (VIS) and the microwave spectra and can be subdivided using different definitions depending on the field of application. Figure 2.1 and Table 2.1 propose one of such classifications based on the atmosphere high-transmission windows as well as the type of detector materials that are sensible on each particular band. Four IR spectral bands are of interest for infrared vision applications: (1) the near infrared (NIR, from ~ 0.7 to 1 µm), (2) the short-wave IR (SWIR, from ~ 1 to 2.5 µm), (3) the mid-wave infrared (MWIR, from ~ 3 to 5 µm), and (4) the long-wave infrared (LWIR, from ~ 7.5 to 14 µm).

Strictly speaking, the terahertz region, which is between 100 and 1000 µm (0.3–30 THz), is part of the very long-wave infrared (VLWIR, from 14 to 1000 µm). However, terahertz radiation can be considered as a special case of thermal infrared radiation, sharing some properties with the infrared and others from the microwave bands.
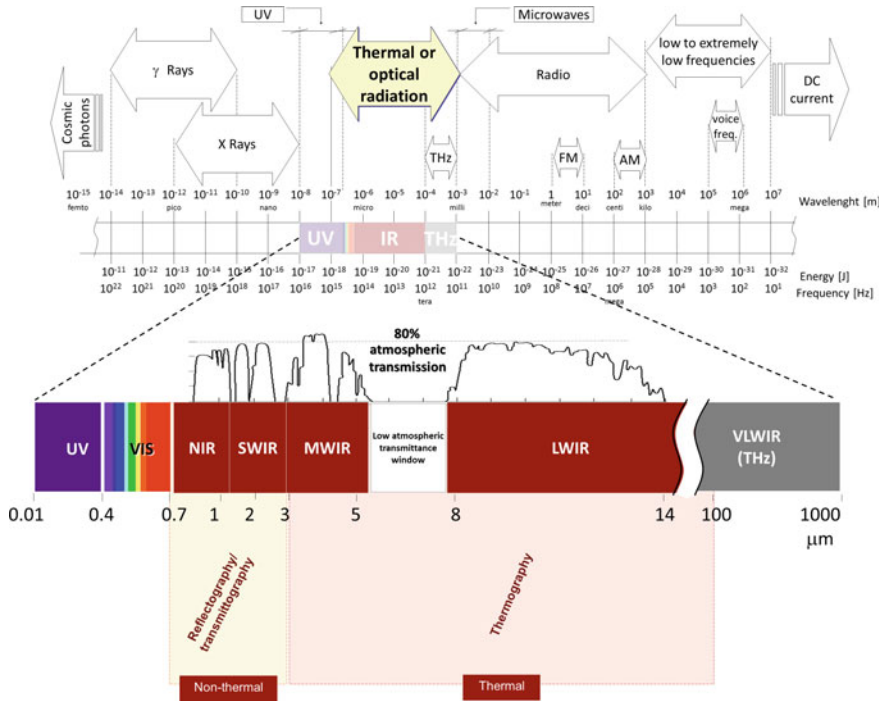
Fig. 2.1 The infrared bands in the electromagnetic spectrum

Table 2.1 Infrared spectral bands and the corresponding typical sensor materials [2, 3]

	Spectral band			Typical sensor materials
Nonthermal	NIR	Near infrared	0.7–1 μm	Si (0.5–1.1 μm)
				InGaAs (0.8–1.8 μm)
	SWIR	Short-wave infrared	1–2.5 μm	PbS (1.3–3.5 μm)
				InGaAs (0.8–1.8 μm)
				InAs (1–3.8 μm)
Thermal	MWIR	Mid-wave infrared	3–5 μm	InSb (3–5.5 μm)
				PbSe (2–6 μm)
				HgCdTe (2–14 μm)
	LWIR	Long-wave infrared	7.5–14 μm	Microbolometers (amorphous silicon, vanadium oxide, etc.)
				HgCdTe (2–14 μm)

An additional and important distinction can be made between the spectral bands presented in Fig. 2.1 and Table 2.1. On the one hand, radiation from the NIR/SWIR spectra is the result of reflections from objects, which is similar to radiations in the visible spectrum, i.e., light, although NIR/SWIR light is not visible to humans, and can be considered as a nonthermal form of infrared radiation. On the other hand,

radiation from the MWIR/LWIR bands originates from the object's surface thermal emissions, and not reflections. These differences between IR bands give rise to a second subdivision of the IR spectrum into two groups: nonthermal and thermal.

2.2   NonThermal Infrared Vision

Nonthermal infrared vision is based on the detection of NIR or SWIR radiation reflected from (reflectography) or transmitted through (transmittography) the object of interest. Infrared reflectography and transmittography are based on the fact that, when an object is exposed to electromagnetic radiation, part of this radiation is reflected from the object surface, one part is transmitted through the object, and another is absorbed by the object. Radiation from the NIR and SWIR bands can thus penetrate some materials and be reflected by others.

This is illustrated in Fig. 2.2 for a semitransparent object without defects (top) and with an internal defect (bottom). The object is irradiated (*I*) from the left side, and as can be observed, part of the radiation passes through the object, another part is reflected, and another is absorbed. In the case of an object with an internal defect, opaque to this kind of radiation, part of the radiation that passes through the semitransparent material will be absorbed and reflected from (but not transmitted through) the defect. This will be revealed in the front (reflectography) and back (transmittography) surfaces of the object.

Proper selection of a continuous and uniform active illumination source is a critical part of a nonthermal inspection system. For instance, incandescent lamps provide a wide electromagnetic spectrum, going from the ultraviolet to the VLWIR (see Fig. 2.1 for reference). Fortunately, a vast part of the radiation from such a source is in the visible and the NIR and SWIR spectral bands, and therefore, it can be used as an illumination source. On the contrary, the EM spectrum of fluorescent lamps is narrower and with only a few distinctive high-intensity peaks mostly in the visible spectrum. Radiation emitted by this kind of lamps in the NIR/SWIR spectra is very limited and can hardly be used as a reliable illumination source. Light-emitting diodes (LEDs) are an example of a very interesting illumination source since they provide a narrow IR spectrum around specific wavelengths, from UV to VLWIR including the NIR and SWIR bands. The radiation source can be combined with the utilization of narrowband filters to further improve contrast [4].

2.3   Thermal Infrared Vision

Thermal infrared vision, also known as infrared thermography [5], is based on the utilization of a MWIR or LWIR infrared sensors, such as infrared cameras, to acquire thermal images or *thermograms*, from the surface of the object being observed. This can be performed in an *active* or a *passive* manner, i.e., with or
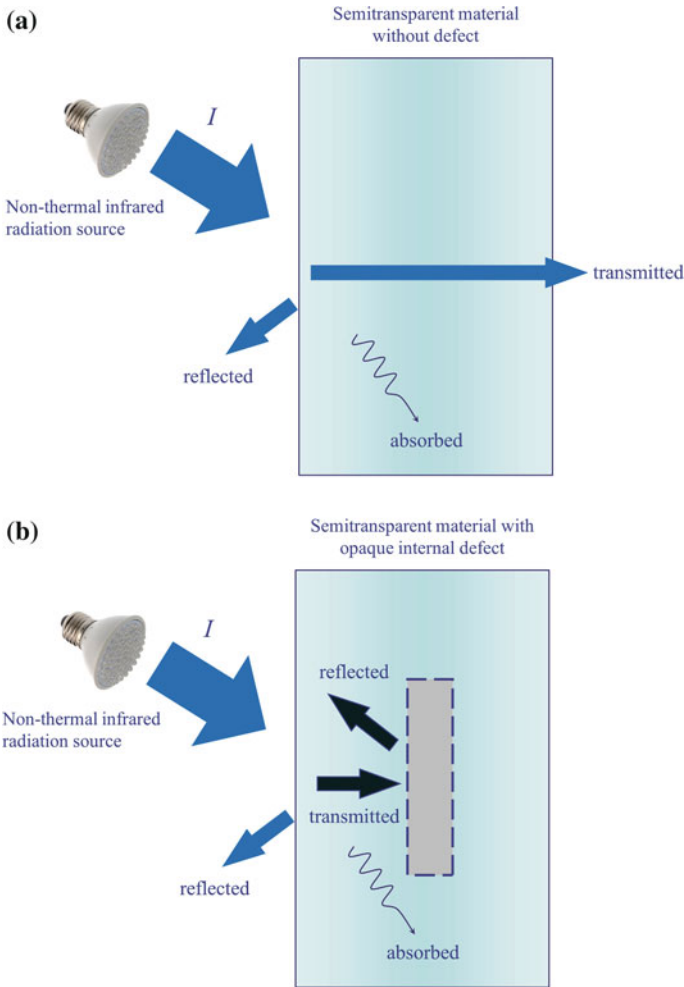
Fig. 2.2 Basic principle behind infrared reflectography/transmittography: semitransparent material exposed to nonthermal infrared radiation **a** without defect. **b** with opaque internal defect

without using an external heating (or cooling) source to promote a thermal contrast between dissimilar materials. In general, active thermography is of greater interest than passive thermography for the nondestructive evaluation (NDE) of materials, since an external thermal stimulation is normally required to disrupt the thermal equilibrium between defective and nondefective areas. Nonetheless, the passive approach has been successfully employed in a number of applications, e.g, aircraft inspection right after landing for water ingress assessment. Active infrared thermography is a well-established NDE technique [5, 6] that allows a fast inspection of large surfaces and is gaining popularity for the inspection of materials.
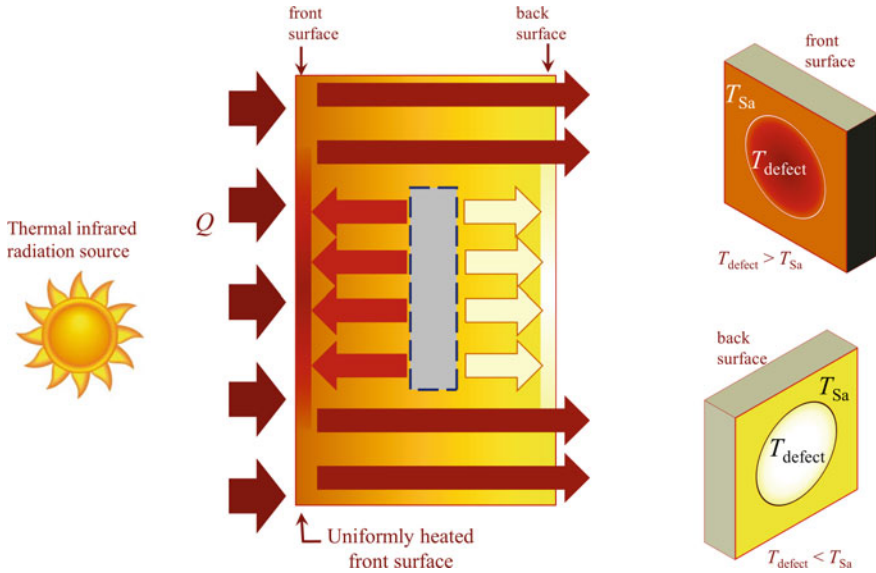
Fig. 2.3 Basic principle behind infrared thermography for an object with an internal defect acting as a thermal resistance

The basic principle behind active infrared thermography is depicted in Fig. 2.3, which illustrates the differences in heat propagation between areas with and without defect, for the case of a square specimen with an internal circular defect. Heat is uniformly applied to the left side surface, after which it propagates through the material by conduction. The heat front is uniformly distributed in the absence of internal defects. On the contrary, if an internal defect is present, it acts as a resistance to heat propagation. This is reflected in both the front and the back surfaces. In this example, it is assumed that the defect acts as a thermal resistance to heat. The situation would be different (the opposite) for the case of a defect acting as a heat sink.

Contrary to infrared reflectography/transmittography, where usually a single image is acquired and processed (after carefully selecting the appropriate illumination source and/or narrowband filters), infrared thermography for NDT applications requires the acquisition of a sequence of thermograms that are processed in time and space.

Several active thermography techniques have been proposed using specific scenarios adapted to particular applications. It is customary to categorize techniques according to the type of energy source and the waveform being employed. For instance, the energy source is commonly classified as follows[5–11]: optical, using either photographic flashes, halogen lamps, infrared lamps, lasers, or any other heat source based on the light emission process; ultrasound [12–14], in which a high-frequency sonic or ultrasonic source (typically 15–40 kHz) is applied to the sample; and inductive [15, 16], based on the principle of heat generation through

eddy currents. On the other hand, the energy delivered to the object can be of different forms: a pulse (few milliseconds), a square pulse or a step (several seconds or minutes), an amplitude and/or frequency modulated wave, or a dynamic scan (point, line, or surface).

Table 2.2 presents a summary of some of the most common active thermography configurations and their designations. According to Table 2.2, an experiment carried out using photographic flashes (short pulse) can be referred as optical pulsed thermography, while the inspection of a part using modulated ultrasounds would be known as ultrasound lock-in thermography.

Table 2.2 Active thermography common technique designations based on the type of energy source and the employed waveform

Source type	Waveform	Technique	Alternative designations
Optical	Short pulse	Optical pulsed thermography	Pulsed thermography (PT)
			Time-resolved thermography (TRIRT)
			Flash thermography
	Long pulse	Optical square pulsed thermography	Square pulsed thermography (SPT)
			Long pulsed thermography
	Step	Optical step heating thermography	Step heating thermography (SHT)
	Amplitude modulated	Optical lock-in thermography	Modulated thermography
	Frequency modulated	Optical chirp thermography	Frequency modulated thermography
	Line scan	Optical line scan thermography	Line scan thermography
Ultrasound	Burst	Ultrasound burst thermography	Vibrothermography (VT)
			Thermosonics
	Amplitude modulated	Ultrasound lock-in thermography	Lock-in vibrothermography
Inductive	Burst	Inductive burst thermography	Eddy current thermography (ECT)
			Pulsed eddy current thermography
	Amplitude modulated	Inductive lock-in thermography	Lock-in eddy current thermography
	Line scan	Inductive line scan thermography	–

2.4 Advanced Signal Processing

Typically, the acquired infrared data are processed to improve defect contrast and signal-to-noise ratio. Several processing techniques exist [5, 11, 17–19], from a basic cold image subtraction to more advanced techniques such as partial least squares thermography (PLST) [17].

The most relevant to the examples presented further are as follows: principal component thermography (PCT) [18], which reorganizes data into new components that take into account the main spatiotemporal variances of the sequence; and higher order statistics thermography (HOST) [19], which calculates the higher order centralized moments (3rd or skewness, 4th or kurtosis, or nth order moment) of the temporal temperature profiles summarizing all the relevant information about the original sequence into a single image.

In addition, in the reflectographic field, the image subtraction between the reflectogram and transmittogram acquired on the same specimen has been previously introduced [20]. The method intensifies the structure of the internal texture of semitransparent materials and shows with a good contrast the damage that has been produced by an impact loading.

For simplicity, only a brief review of the techniques has been provided. More details can be consulted in the provided references.

2.5 Experimental Configuration

The experimental configuration for nonthermal and thermal infrared vision is illustrated in Fig. 2.4. The radiation source ❶, thermal or nonthermal, is pointed toward the inspected object ❷ in either reflection (same side as the camera) or transmission (camera on the opposite side). The camera ❸ (NIR, SWIR, MWIR, or LWIR) records images from the object surface. The image acquisition and radiation sources are synchronized and controlled by a computer or control card ❹. Data (single images or sequences of several images) are stored and processed with a computer ❺.

Figure 2.4 is a schematic of the IR vision experimental setup using an optical source, i.e, an incandescent light bulb, which delivers wide-spectrum IR radiation covering both nonthermal and thermal applications. In practice, the illumination or heat sources are carefully selected depending on the application.
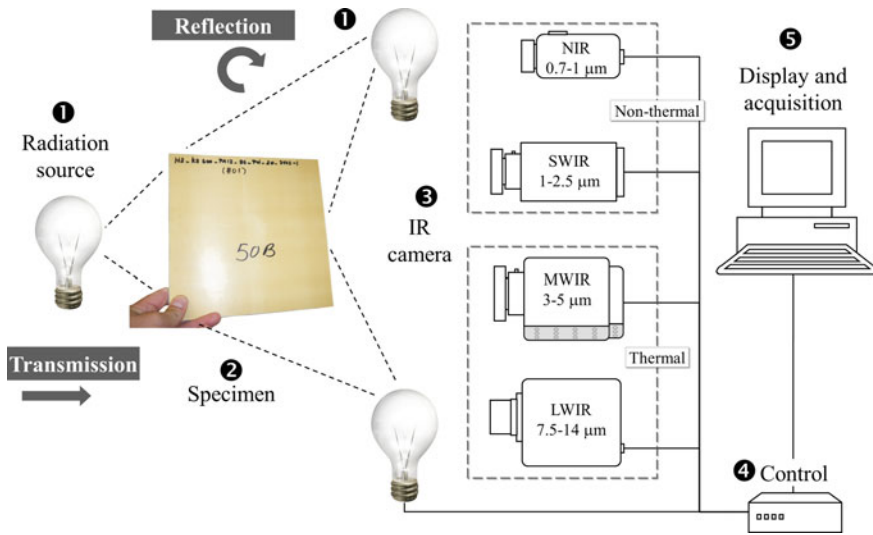
Fig. 2.4 Typical experimental setup for infrared vision

2.6 Integrated Imaging Applications

2.6.1 Inspection of Aramid–Phenolic Composites by Near Infrared Reflectography/Transmittography

The use of NIR reflectography/transmittography has shown to be very promising for the assessment of composite materials' semitransparent in the NIR/SWIR spectra [21]. For instance, Fig. 2.5 presents some results from an aramid–phenolic composite impacted specimen inspected from both sides in reflection and transmission modes. The top row corresponds to the front face and the bottom row to the back face of the specimen.

Photographs in Fig. 2.5a show the front (top) and back (bottom) views of the specimen. Some subtle indications of the impact can be observed in the close-up views in Fig. 2.5b, corresponding to approximately the area delimited by the black rectangle in Fig. 2.5a. The SWIR reflectograms presented in Fig. 2.5c show more details about the damage caused by the impact in both faces. Among other things, the specimen presents folds in three cardinal points around the impact (north, east, and west) as observed from the front face reflectogram (Fig. 2.5c, top). The back face reflectogram (Fig. 2.5c, bottom) only shows a fold north of the impact. However, contrary to front face inspection (Fig. 2.5c, top) where single fold is observed, the presence of two folds is evidenced in the back face inspection (Fig. 2.5c, bottom).

The transmittograms presented in Fig. 2.5d confirm the presence of the three folds visible from front face reflectogram (Fig. 2.5c, top) and apparently a fourth

Front

Back

(a)   (b)   (c)   (d)
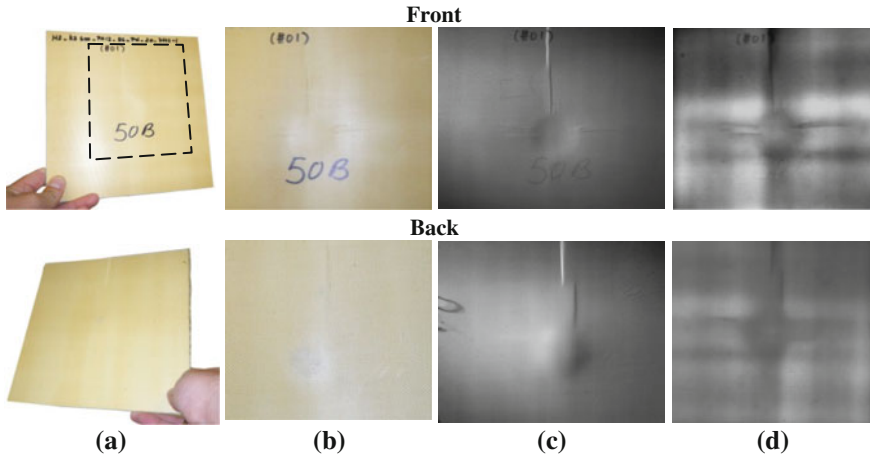
Fig. 2.5 Aramid–phenolic composite impacted specimen, front (*top row*) and back (*bottom row*) faces: **a** specimen photographs; **b** close-up views of the impacted area; **c** NIR reflectograms at 940 nm; and **d** NIR transmittograms at 780 nm

one south of the impact. Additionally, Fig. 2.5d provides information about the fibers and epoxy distribution inside the specimen. There are two bright stripes, one north and one south of the impact, indicating that fiber content is lesser in these areas, probably as a consequence of the impact that forced the fiber in the vertical and horizontal to congregate at the center of the impact.

2.6.2 Integrated Infrared Vision Inspection of Jute and Hemp Fibers

An example showing the complementarity of infrared transmittography and infrared thermography is presented in Fig. 2.6. These images correspond to a natural fiber composite specimen manufactured using six inner layers of jute plain weave disposed at 0°/0°/45°/°– 45°/0°/0° sandwiched between two nonwoven hemp mats. The laminate in this example has been impacted with a force 2339 N. Figure 2.6a shows the front side and Fig. 2.6b the back side of the specimen.

The specimen was inspected from the back side (radiated from the front side) by NIR and SWIR transmittography, and the results are shown in Fig. 2.6c, d, respectively. It is possible to see the extent of the damage inside the sample as well as the fiber distribution in both cases, although with some differences in contrast.

The sample was also inspected by square pulsed thermography (SPT) [22], 4 min heating–9 min cooling, using one 500-W lamp positioned in reflection mode from the back side. Data were then processed by principal component thermography (PCT) [23]. The resulting images correspond to the empirical orthogonal
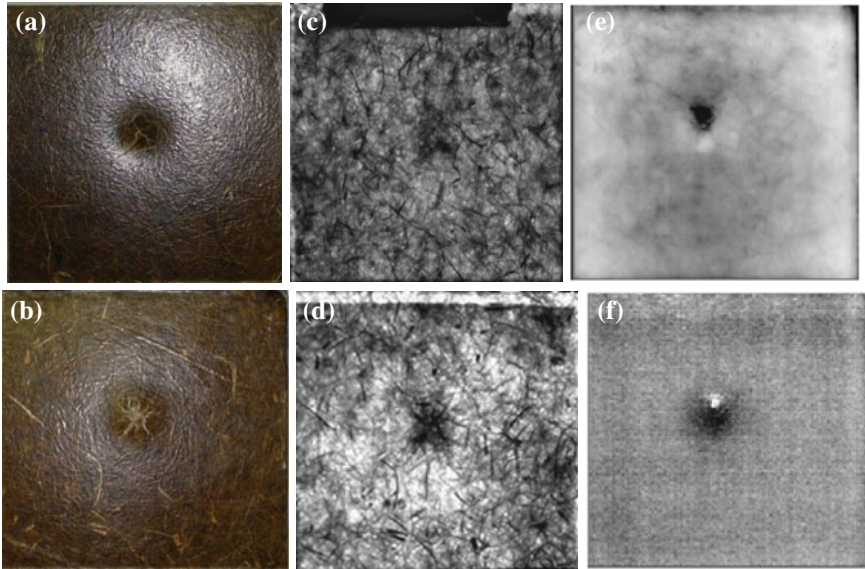
Fig. 2.6 **a** Front and **b** rear views of the specimen; **c** NIR (0.9–1.0 nm) transmittogram from the front face, **d** SWIR (1430 nm) transmittogram from the rear face; PCT results from square pulsed thermography (reflection from the front face): **e** EOF_3 and **f** EOF_{12}

functions (EOF), each one showing information at different depths. The EOF_3 in Fig. 2.6e is related to the damage close to the surface, while the EOF_{12} in Fig. 2.6f corresponds to the damage deeper in the sample. As can be noted, the damaged area grows from the front to the back faces of the specimen: It is smaller in Fig. 2.6e than in Fig. 2.6f.

A similar specimen but impacted with considerably less force a maximum 1843 N is shown in Fig. 2.7.

Front and rear sides are reported in Fig. 2.7a, b, respectively. The transmittograms (Fig. 2.7c, d) show the presence of some areas at higher density of fibers highlighted with dark color, while the white color distinguishes the areas of low fiber density. Also in this case, the growth of the number linked to the EOF (from 4 to 14) indicates a growth of the depth explored starting from the face (Fig. 2.7e) to the rear side (Fig. 2.7f).

2.6.3 Assessment of Pultruded Kenaf/Glass Hybrid Composites

Another interesting case of semitransparent composite material is shown in Fig. 2.8. The specimen contains 40 kenaf + 25 glass as fiber content (wt%) and S/C/K/C/S as stacking sequence, with S = protection surface veil (made of 0.2 mm thickness
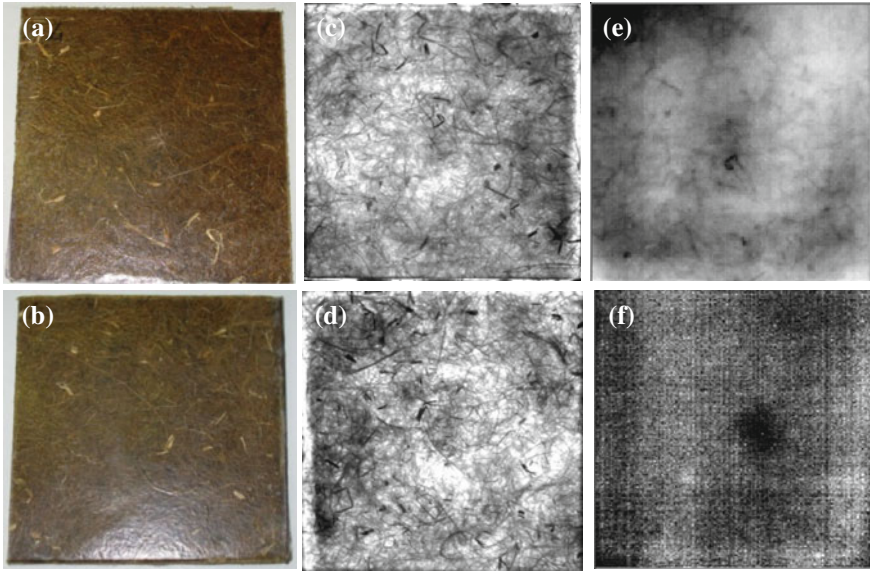
Fig. 2.7 **a** Front and **b** rear views of the specimen; **c** SWIR (1300 nm) transmittogram from the front face, **d** SWIR (1300 nm) transmittogram from the rear face; PCT results from square pulsed thermography (reflection from the front face): **e** EOF_4 and **f** EOF_{14}

polyester fibers); C = chopped strand mat (glass fibers); and K = kenaf fibers. The final dimensions were 140 mm x 22 mm x 2.8 ($\pm$ 0.1) mm.

Indentation test was carried out by supporting the specimen on a steel plate with a 30 mm diameter circular opening, using a 10-mm-diameter hemispherical indenter as well as a 5584 universal testing machine (Instron) in displacement control with 1 mm/min crosshead speed [24].

As seen in Fig. 2.8, the internal structure of the composite is very different if compared to the external layer. The visible image in Fig. 2.8a shows a nonhomogeneous texture of the front side. For example, there is a horizontal rectangular area marked by an arrow that is also identifiable in the rear side (Fig. 2.8b). The energy applied in the indentation test has almost broken down the specimen (Fig. 2.8b—magnifications).

Comparing the reflectogram in Fig. 2.8c to the visible image of the same side (Fig. 2.8b), it is possible to notice that the described area disappears and the alignment of the external fibers is enhanced for the whole specimen.

Working in transmission mode (Fig. 2.8d), the specimen is discriminated like a radiograph, i.e., the areas having a higher concentration of fibers are distinguished from the others. However, a more in-depth study about this point is only possible by subtracting the reflectogram (Fig. 2.8c) to the transmittogram (Fig. 2.8d). Indeed, some areas classified at low concentration of fibers in the first approach are effectively more saturated at the end of the second step, as highlighted by the arrows in Fig. 2.8d, e.
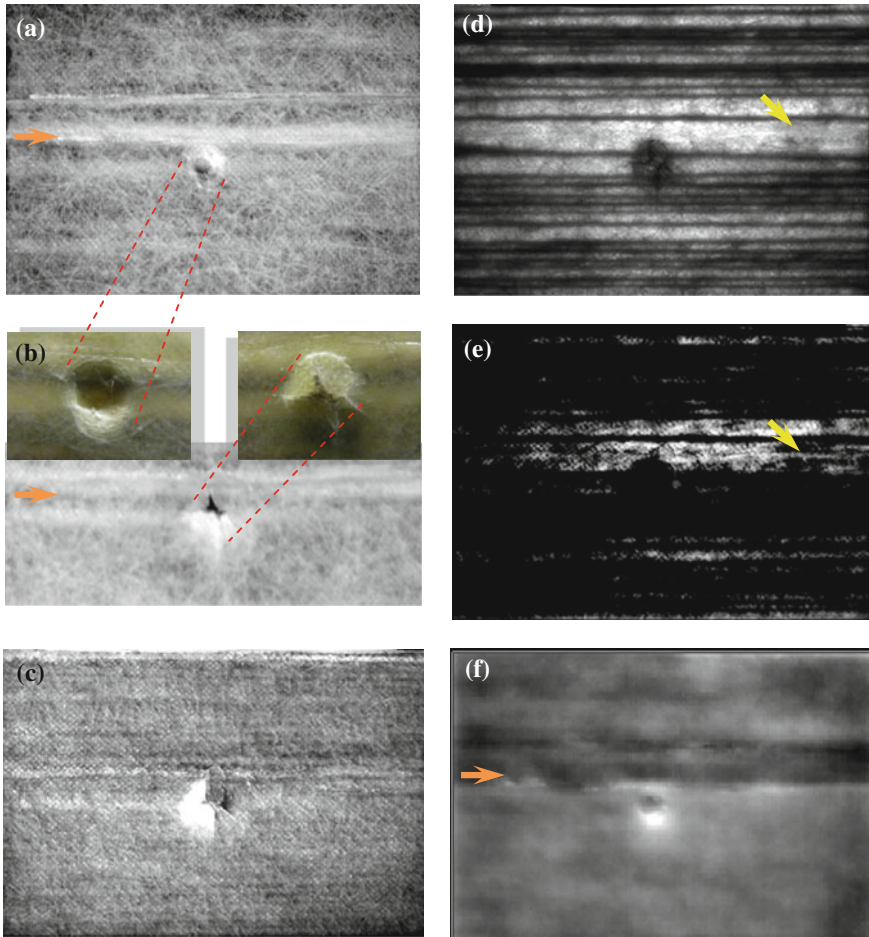
Fig. 2.8 **a** Front and **b** rear (close-up) views of the specimen and magnifications of the impact damages; **c** SWIR (940 nm) reflectogram from the rear face; **d** SWIR (940 nm) transmittogram from the rear face, **e** subtraction of images between **c** and **d**; PCT result from square pulsed thermography (reflection from the front face); and **f** EOF$_3$

Finally, the PCT result (EOF$_3$) presented in Fig. 2.8f reveals a horizontal rectangular area that has a larger extension compared to that identified by visual inspection in Fig. 2.8a, demonstrating the potential of thermography to identify subsurface features related with fiber distribution.

2.6.4 Impacted Glass and Basalt Fiber Woven Composites

Another example of integrated approach between visual inspection, SWIR, and thermographic techniques is presented in Fig. 2.9. The composite under inspection was produced by basalt fibers. The fiber volume fraction was equal to 0.38 ± 0.02, and the thickness was approximately equal to 3 ± 0.1 mm [25]. Low-velocity impact test was conducted using a drop-weight impact tower fitted with an anti-rebound device. The impactor, with a 12.7 mm diameter, consists of a piezoelectric load cell to record the force and a high-speed video camera to record the impactor displacement during the impact test. Prior to undergoing impact loading, the sample was clamped to a square frame with a 73-mm square window. The mass of the bare impact carriage was 5 kg. The impact energy was equal to 15 J, and it was obtained by setting the impact height at 300 mm [26].

The front and the back sides of the specimen are shown in Fig. 2.9a, b, respectively. The damage caused by the impact is partially visible to the naked eye.

A very large detachment appears using SWIR when working in reflection mode, although at low contrast (Fig. 2.9c). The same defect can be fully detected in transmission mode by SWIR at 1050 nm (Fig. 2.9d); at the same time, the damage due to the impact loading can be revealed. In addition, the latter method enhances the visualization of a different texture of the fibers, compared to the surrounding area, where the dotted rectangle is positioned.
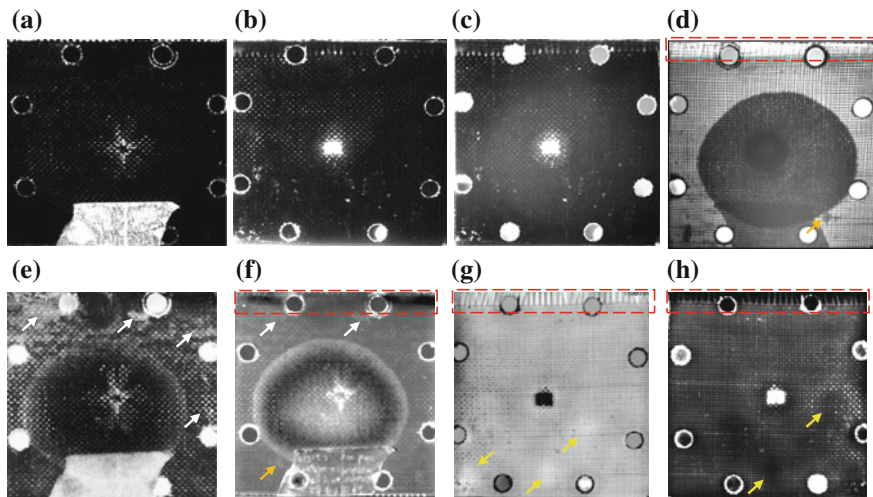
Fig. 2.9 **a** Front and **b** rear views of the specimen; **c** SWIR (1050 nm) reflectogram from the rear face, **d** SWIR (1050 nm) transmittogram from the back face; **e** HOST result (skewness) from pulsed thermography (reflection from the front face), **f** PCT result (EOF_4) from pulsed thermography (reflection from the front face), **g** PCT result (EOF_4) from pulsed thermography (reflection from the back face), and **h** HOST result (skewness) from pulsed thermography (reflection from the back face)

Some micro-detachments, pointed out by arrows, can be detected working in PT configuration and analyzing the IR data with HOST (skewness). Using this type of processing, the contour of the large detachment that surrounds the trace of the impact is still identifiable, while no thermal anomaly linked to statistical parameters reveals the presence of fiber misalignments at the north side of the specimen (Fig. 2.9e).

The latter defect is mapped by a dark color in Fig. 2.9f, by processing the IR data with PCT technique. In the same figure, some micro-detachments previously described are visualized and marked by arrows. One of these was already signaled in Fig. 2.9d by using an orange–brown arrow.

Additional anomalies are detected, identified by arrows in Fig. 2.9g, after inspecting the specimen from the back face, which also shows the texture of the fibers on the whole specimen. Finally, the detachments that are located below the large central detachment (Fig. 2.9g) are confirmed using HOST technique (skewness). The same consideration is valid for the fiber misalignments located along the north side of the specimen (Fig. 2.9h).

The large detachment around and above the impact could be due to poor adhesion between the resin and the fibers during sample fabrication.

2.6.5   Integrated Inspection of Jute Cloth/Wool Felts Hybrid Laminates

In Fig. 2.10 is shown a wool-based composite containing an amount of 40 vol.% (W-40) of wool fiber. From this laminate was cut the specimen. The specimen of dimensions 170 mm × 25 mm × 4 mm ($L \times W \times t$) was subject to a low-velocity impact test using an in-house built drop-weight impact tower. The impactor diameter was 12.7 mm, and its mass was 1.25 kg, which was dropped from a height of 0.5 m, for an impact energy of approximately 6 J.

The specimen was produced by hand layup technique in a closed mold of dimensions 260 × 160 mm at room temperature using an epoxy resin (Mates SX10 ver.2 with hardener SX14). The laminate was cured under a slight pressure of 0.02 bars for 2 h, and then, it was removed from the mold and further cured at room temperature for at least 24 h before use [27].

The visual inspection of the front side (Fig. 2.10a) shows the damage produced by the impactor at the center of the specimen, while the rear side displays two dark areas. The first one, surrounded by an oval (Fig. 2.10b), is effectively larger than that visible to the naked eye. This assumption is confirmed in the near IR working in transmission mode (Fig. 2.10c) by using a metal-oxide semiconductor camera (22.2 × 14.8 mm, 10 MP, spectral band between 0.38 and 1.0 μm), with a visible cutoff filter to limit the spectrum from 0.7 to 1.0 μm.

The second one, surrounded by a square, can be seen with a good contrast in the near IR working in reflection mode (Fig. 2.10d).
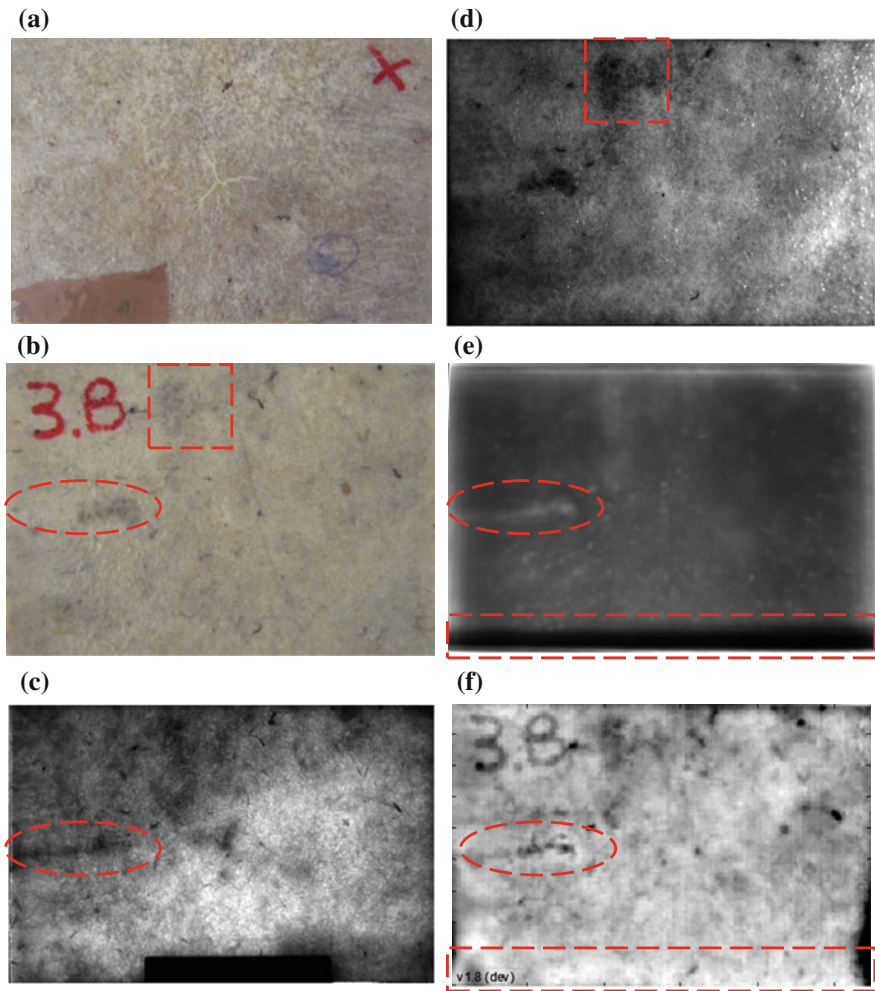
Fig. 2.10 **a** Front and **b** rear views of the specimen; **c** NIR (CMOS camera) transmittogram from the rear face, **d** NIR (CMOS camera) reflectogram from the rear face; **e** PCT result (EOF$_2$) from square pulsed thermography (reflection from the rear face), **f** PCT result (EOF$_2$) from pulsed thermography (reflection from the rear face)

In addition, the first anomaly (partially superficial and partially sub-superficial) is detected entirely by using square pulsed thermography (SPT) (Fig. 2.10e) and pulsed thermography (PT) (Fig. 2.10f).

PCT technique is able to reveal the presence of a horizontal rectangular area positioned along a border of the specimen, but with a reverse color, i.e., dark analyzing the SPT data (Fig. 2.10e) and white analyzing the PT data (Fig. 2.10f).

2.7 Conclusions

Infrared vision, thermal and nonthermal, is a very promising technique that is still finding new and interesting applications. NIR/SWIR reflectography/transmittography have shown a great potential for the inspection of semitransparent materials such as composites (glass, aramid–phenolic, or natural fibers). Active thermography is already widely used for the NDT of several materials. The observed complementarity between NIR/SWIR reflectography/transmittography and MWIR/LWIR thermography opens the door for integrated solutions for specific applications.

References

1. Gonzalez RC, Woods RE, Eddins SL (2004) Digital image processing using MATLAB. Pearson Prentice Hall, Upper Saddle River, NJ
2. Rogalski A (2011) Recent progress in infrared detector technologies. Infrared Phys Technol 54:136–154
3. Rogalski A, Chrzanowski K (2002) Infrared devices and techniques. Opto-Electron Rev 10 (2):111–136
4. Ebeid A, Rott S, Talmy E, Ibarra-Castanedo C, Bendada A, Maldague X (2010) Near infrared imaging for multi-polar civilian applications. In: Proceedings of 10th International Conference on Quantitative InfraRed Thermography—QIRT, Quebec City, Canada, 27–30 July 2010
5. Maldague XP (2001) Theory and practice of infrared technology for nondestructive testing. Wiley, New York
6. Nondestructive Handbook. In: Maldague X, Moore PO Infrared and thermal testing, vol 3, 3rd edn. ASNT Press, Columbus, 718 p
7. Vollmer M, Möllmann K-P (2010) Infrared thermal imaging: fundamentals, research and applications. Wiley-VCH, Weinheim Germany
8. MeolaC (ed) (2012) Infrared thermography recent advances and future trends. Bentham eBooks, Sharjah
9. Holst GC (2000) Common sense approach to thermal imaging. SPIE Optical Engineering Press, Bellingham
10. Ibarra-Castanedo C, Tarpani J-R, Maldague XP (2013) Nondestructive testing with thermography. Eur J Phys 34:S91–S109
11. Ibarra-Castanedo C, Genest M, Piau J-M, Guibert S., Bendada A, Maldague XPV (2007) Chapter 16: active infrared thermography techniques for the nondestructive testing of materials. In: Chen CH (ed) Ultrasonic and advanced methods for nondestructive testing and material characterization, World Scientific Publishing, Singapore, 684p
12. Favro LD, Han X, Ouyang Z, Sun G, Sui H, Thomas RL (2000) Infrared imaging of defects heated by a sonic pulse. Rev Sci Instrum 71(6):2418–2421
13. Shepard SM, Ahmed T, Lhota JR (2004) Experimental considerations in vibrothermography. In: Burleigh DD, Cramer KE, Peacock GR (eds) Proceedings of SPIE—the international society for optical engineering, vol 5405. Thermosense XXVI, Orlando, pp 332–335

14. Zweschper T, Riegert G, Dillenz A, Busse G (2005) Ultrasound excited thermography—advances due to frequency modulated elastic waves. J. QIRT 2(1):65–76
15. Riegert G, Zweschper T, Busse G (2004) Lock-in thermography with eddy-current excitation. J. QIRT 1(1):21–32
16. Grenier M, Ibarra-Castanedo C, Luneau F Bendada H, Maldague X (2002) Development of a field concentrator coil by finite element modeling for power efficiency optimization in eddy current thermography inspection. In: Thompson DO, Chimenti DE (eds) Review of quantitative nondestructive evaluation vol 21, pp 3–13
17. Lopez F, Nicolau V, Maldague X, Ibarra-Castanedo C (2013) Multivariate infrared signal processing by partial least-squares thermography. In: international symposium on applied electromagnetics and mechanics. Québec City, Canada, 31 July 2013–2 Aug 2013
18. Rajic N (2002) Principal component thermography, DSTO-TR-1298, http://www.dsto.defence.gov.au/publications/2452/DSTO-TR-1298.pdf. Accessed on 19 Dec 2013
19. Madruga FJ, Ibarra-Castanedo C, Conde OM, Lopez-Higuera J, Maldague X (2010) Infrared thermography based on higher-order statistics. NDT&E Int 43(8):661–666
20. Sfarra S, Ibarra-Castanedo C, Santulli C, Sarasini F, Ambrosini D, Paoletti D, Maldague X (2013) Eco-friendly laminates: from the indentation to non-destructive evaluation by optical and infrared monitoring techniques. Strain 49:175–189
21. Bendada A, Sfarra S, Genest M, Paoletti D, Rott S, Talmy E, Ibarra-Castanedo C, Maldague X (2013) How to reveal subsurface defects in Kevlar® composite materials after an impact loading using infrared vision and optical NDT techniques? Eng Fract Mech 108:195–208
22. Sfarra S, Bendada A, Paoletti A, Paoletti D, Ambrosini D, Ibarra-Castanedo C, Maldague X (2010) Square pulse thermography (SPT) and digital speckle photography (DSP): non-destructive testing (NDT) techniques applied to the defects detection in aerospace materials. In: International symposium on NDT in aerospace, Hamburg, Germany, 22–24 Nov 2010. Available online: http://www.ndt.net/article/aero2010/papers/p3.pdf
23. Rajic N (2002) Principal component thermography for flaw contrast enhancement and flaw depth characterization in composite structures. Comp Struct 58:521–528
24. Md. Akid H, De Rosa IM, Santulli C, Sarasini F (2010) Flexural behaviour of pultruded jute/glass and kenaf/glass hybrid composites monitored using acoustic emission. Mat Sci Eng A-Struct 527:2942–2950
25. Carmisciano S, De Rosa IM, Sarasini F, Tamburrano A, Valente M (2011) Basalt woven fibre reinforced vinylester composites: flexural and electrical properties. Mater Design 32(1):337–342
26. Sfarra S, Ibarra-Castanedo C, Santulli C, Paoletti A, Paoletti D, Sarasini F, Bendada A, Maldague X (2013) Falling weight impacted glass and basalt fibre woven composites inspected using non-destructive techniques. Compos Part B-Eng 45:601–608
27. Santulli C, Sarasini F, Tirillò J, Valente T, Valente M, Caruso AP, Infantino M, Nisini E, Minak G (2013) Mechanical behaviour of jute cloth/wool felts hybrid laminates. Mater Design 50:309–321

Chapter 3
Inspection Methods for Metal Surfaces: Image Acquisition and Algorithms for the Characterization of Defects

Reinhold Huber-Mörk, Gustavo Fernández Domínguez, Svorad Štolc,
Daniel Soukup and Csaba Beleznai

Abstract This chapter provides an overview of acquisition methods, specific challenges, and solutions for image-based metal surface inspection. We discuss illumination and recording methods for a wide range of tasks such as inspection of rail surfaces, weld quality assessment, inspection of holograms on metallic foils, and reconstruction of coins. Algorithms exploiting the properties of the acquired image data are presented in detail. Image acquisition, tracking, and segmentation of weld beads in harsh environments are presented. Reference and non-reference weld quality assessments are discussed. Detection of surface disruptions as well as surface classification for images of rails from a mobile acquisition platform is demonstrated. Multi-view approaches are explored for the inspection of optically variable security elements on metallic foil and light-field data processing is applied to the inspection of coins.

Contents

R. Huber-Mörk (✉) · G.F. Domínguez · S. Štolc · D. Soukup · C. Beleznai
Intelligent Vision Systems, AIT Austrian Institute of Technology, Vienna, Austria
e-mail: reinhold.huber-moerk@ait.ac.at

3.1   Introduction

Automated visual inspection of metal surfaces often imposes a challenge due to several specific factors. Metal surfaces reflect, scatter, sometimes absorb light, and often exhibit strong directional variation of observation conditions. Furthermore, metal surface characteristics might substantially vary from one material or surface manipulation characteristic to another, such as from specular to diffuse, and might consist of different surface textures and structures: two-dimensional compact features (spots, grains), 2D linear patterns (boundaries, seams, scratches), or volumetric artifacts such as cracks, holes and other spatial deviations from a given reference 3D shape. In addition, these texture and structural patterns often appear differently at various spatial scales and magnifications, thus strongly influencing the optimum choice of imaging hardware. This multitude of factors affecting the inspection process calls for a rigorous and thorough analysis in each specific inspection task, in terms of material properties, illumination conditions, observation geometry, and vision hardware.

Various publications addressing different inspection tasks related to metal surfaces appeared. In order to demonstrate the broad range of applications and approaches, we provide a few of them. Anomaly detection in milled steel surfaces using various approaches for 2D texture analysis was discussed [20]. A method using a series of bright- and dark-field illuminated images and 2.5D analysis was used for solder paste inspection [37]. An experimental study describes specular reflection from metallic surfaces and application to coin classification using images illuminated by light sources differing in location and spectral emission [22]. Detection of 3D surface defects on specular metal sheets using infrared deflectometry (i.e., a reference pattern is projected to the surface, captured by a camera and analyzed) was proposed for applications in the automotive industry [45]. Three-dimensional information is obtained from the heat transfer properties by heating the surface of a metallic industrial part by means of a laser source and image capture in the infrared range [2].

Many attempts have been made to review and categorize the field of automated industrial visual inspection (AVI). Although those studies tried to review the different tasks as well as the diverse solutions as much as possible, specific questions are not addressed. In this chapter, we restricted ourselves to vision in the visible spectrum (with respect to human vision capabilities) and real-time (or close to it) computational operation applied to issues occurring in metallic surface inspection.

In detail, this chapter presents a number of representative applied examples where many challenges specific to metal surfaces are addressed and operational vision system solutions are outlined. We describe image acquisition, artifact

detection, and classification for specific inspection tasks. Acquisition setups inspired by photometric and conventional stereo imaging applicable to mobile platform and industrial line-scan camera configurations are introduced. An approach for detection of deviations from expected surface shape is introduced, and a method for classification of surface properties is discussed.

An automated welding seam inspection framework is presented assessing the seam quality (width, texture variations, and surface characteristics) during arc welding in an online manner. Given the highly dynamic nature of the observation conditions (flashing lights, sparks, smoke, and moving metal workpiece), the system employs a specific vision setup with stroboscopic lighting and an algorithmic chain based on robust statistics and machine learning in order to cope with the difficult conditions.

Another 2.5D acquisition setup consisting of a color line-scan camera and two light sources operating in different spectral bands is applied to acquire images of rails from a mobile platform. The most severe defects observed on rails are small cavities in the surface. Additionally, the characteristics of surface regions which could be related to surface patterns and roughness are of interest. Analysis of local spectral properties is employed for the detection of small defects. We compare approaches analyzing the anti-correlation and co-occurrence properties between spectral channels in order to characterize local deviations from an expected model of undistorted surface properties.

In order to complement 2.5D analysis, we introduce a 3D approach derived from traditional stereo imaging. In particular, we demonstrate a line-scan multi-view stereo system using a single area-scan sensor, a single lens, and planar mirrors attached to the acquisition device. In this case, the inspected object is moving relatively to the acquisition device, where each acquired line is always observed under three different angles at the same time. The relative motion between object and acquisition device inherently fulfills the epipolar constraint in stereo vision.

For the analysis of surface region properties, we relate the reflectance properties of the surface patterns and roughness to image texture. Responses of the Gabor filter bank are used as textural features. Texture parameters corresponding to relevant surface structure classes are modeled by Gaussian mixtures and fused with background knowledge in a Bayesian classifier. Regions dominated by structures such as low distortion, corrosion, microcracks, and grinding could be identified by the texture analysis. Fusion with defects detected by 2.5D analysis results in a map of the rail condition and wear.

Finally, we provide alternative approaches for multi-view metal surface inspection which promises improved accuracy at smaller scales, however, are much more challenging in terms of a data acquisition and processing. In particular, we introduce a stereo system and a line-scan light-field approach which observes an object under a plurality of views. The stereo system obtains depth and the light-field system additionally obtains an enhanced depth of field. The objects under investigation are metallic security features and coins, respectively.

The chapter attempts to provide an in-depth look of the presented inspection systems and demonstrates numerous results of substantial practical impact.

3.2 Image Acquisition for Metallic Surfaces

The appearance of metal surfaces in machine vision depends both on material properties, and illumination and acquisition geometry. Appropriate relative placement of illumination sources, cameras, and observed objects enhances specific object properties, e.g., texture, edges, and surface deformations, and enables automatic analysis methods. The chosen acquisition setup is determined from the goal of the analysis, e.g., defect detection, surface reconstruction, and material classification, and optical properties of the object under observation, e.g., specular and diffuse reflection, refraction, and absorption. We start reviewing basic principles governing the appearance of images of metal surfaces and subsequently give an overview of acquisition setups in machine vision.

3.2.1 Material Surface Properties and Reflection

For metal surfaces, the relationship between specular reflection and diffuse scattering depends primarily on the light wavelength and surface roughness. In general, the extent of diffuse scattering does not only depend on surface roughness, it rather stems from subsurface interaction of light and material. For metals, the influence of subsurface scattering can be neglected most of the time as visible light does not penetrate beyond the surface.

Surface roughness is commonly either defined by the mean absolute R_a or the root mean squared (RMS) R_q deviation from the average height of a surface patch characterized by elevation measurements $z_i, i = 1, \ldots, n$

$$R_a = \frac{1}{n} \sum_{i=1}^{n} |z_i - \bar{z}| , \quad R_q = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (z_i - \bar{z})^2} , \quad \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i . \quad (3.1)$$

The relationship between surface roughness and optical scattering is given by the total integrated scatter (TIS) introduced by Bennett and Porteus [5]

$$\text{TIS} = R_0 \left(1 - \exp\left(-\frac{4\pi R_q \cos \theta}{\lambda} \right) \right), \quad (3.2)$$

where R_0 is the reflectance of a perfectly smooth surface of the considered material, θ is the angle of incidence, and λ is the wavelength of the light. For metallic surfaces, we set $R_0 = 1$, which refers to ideal specular and can be assumed for $R_a \ll \lambda$. On the left of Fig. 3.1, the dependency of TIS on different incidence angles for a number of surface roughness R_q values (a $\lambda = 0.5\,\mu\text{m}$ was assumed) is shown. Ideal specular reflection depending on the material is described by the Fresnel reflectance which shows little variation depending on θ for metals and is
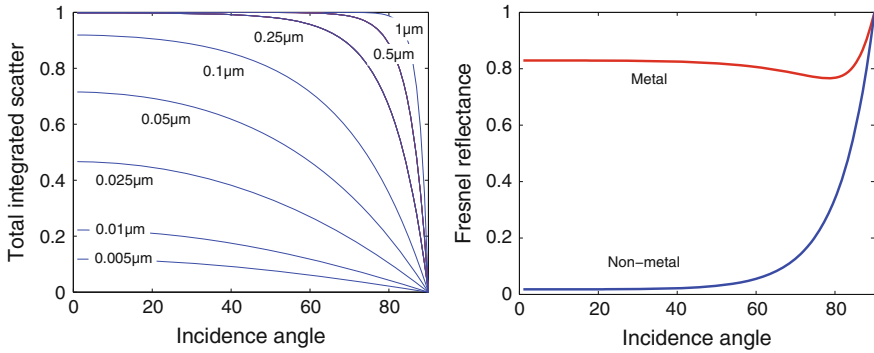
Fig. 3.1 Scattering from surfaces depending on angle of incidence θ: Total integrated scatter given for different surface roughness values R_q with $\lambda = 0.5\,\mu m$ (*Left*), Fresnel reflectance for typical metal and nonmetal surface

typically very high when compared to nonmetals, see Fig. 3.1 on the right. More details on reflectance properties and modeling can be found in [61].

For precisely polished metal surfaces, we could expect R_q on the order of magnitude of $0.005\,\mu m$ and a TIS ≈ 0 (assuming an angle of incidence of $\theta = 45°$ and a wavelength of $\lambda = 0.5\,\mu m$). In such a case, the amount of reflected light predominantly depends on the placement of light source and observer given by the law of reflection (incoming angle equals outgoing angle with respect to surface normal). For grinded metal surfaces, a R_q on the order of magnitude of $0.5\,\mu m$, a TIS ≈ 1 is obtained using the same settings as above. We will employ those reflection properties in machine vision for surface analysis.

3.2.2 Acquisition System Properties

In their "Survey of Automated Visual Inspection" (AVI), Newman and Jain point out the diversity of demands on AVI systems and the need to fit such systems to the individual needs of each environment [34]. In the following section, we will discuss machine vision setups, i.e., selection and placing of cameras and lighting, tailored to metal surface inspection in industrial environments.

Concerning cameras we restrict ourself to cameras sensitive in the visual range, i.e., roughly $380, \ldots, 780\,nm$. Typical industrial settings involve relative movement between observed object and observing hardware, e.g., a product transported on a conveyor belt or a recording device moving on a mobile platform. A line-scan camera synchronized with this relative motion offers continuous images, less distortion, lower requirements on lighting, and higher sensitivity, especially when "time delay and integration" (TDI) technology is used [63]. In TDI, each object line is observed multiple times while sensor and object move relatively with respect to each other. Hence, a number of sensor lines are recorded for each object line from

the same viewpoint (at different time instances). Summation of responses for corresponding sensor lines delivers a single object line with increased sensitivity.

Acquisition of several images or lines of the same object from different viewing positions implements a multi-view camera system. Multi-view imaging enables 2.5D and 3D approaches for reconstruction and analysis. The term 2.5D is loosely defined, and we will call any setup a 2.5D system where a point is characterized by two planar coordinates plus an additional property, e.g., ranging from the deviation from a flat plane to measured depth. Traditional stereo imaging as well as more elaborated light-field camera setups is possible.

In machine vision, one usually controls the illumination conditions by task-dependent placement of light sources. The largest difference is between bright- and dark-field illumination. While in bright-field illumination, the goal is to direct the illumination from the observed object via reflection to the sensor; reflection from the object toward the sensor is avoided in a dark-field setup. Figure 3.2 shows light placement positions for a camera looking top-down onto a specular surface. Certainly, reflection toward the sensor occurs in dark field as well, but it is typically restricted to discontinuities, edges, and distortions, which is in turn used in analysis. Figure 3.3 shows images of the same coin under bright- and dark-field illumination.

Active illumination offers a variety of possibilities with respect to geometrical arrangement, spectral properties, and structuring of light (e.g., pattern projection). Geometrical arrangement of light sources enables 2.5D–3D analysis approaches such as shape from shading [64] and photometric stereo [4], where images under different illumination are used to reconstruct the surface under investigation or, at least, to infer surface characteristics. More elaborated illumination setups include projection of simple structured light patterns, e.g., a laser line in laser triangulation, or even a more complicated pattern as also done in commercial applications, e.g., projection of random patterns in the infrared range in the first generation of the Microsoft Kinect sensor [68].

Fig. 3.2 Camera and light placement for *bright-* and *dark-field* acquisition of a specular surface
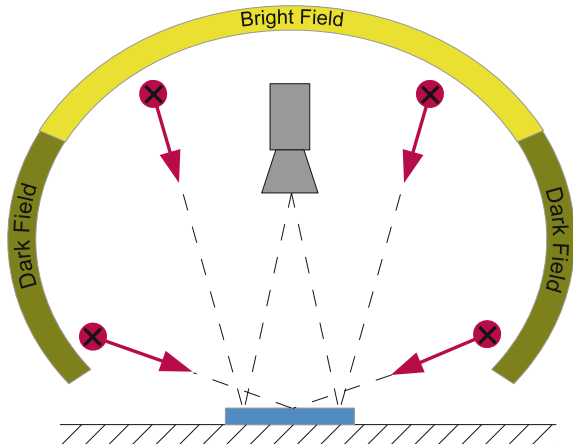
Fig. 3.3 Images of a structured metallic surface under *bright-field* illumination (*left*) and *dark-field* illumination (*right*)

3.3 Weld Quality Assessment

Arc welding has been known for about 200 years, but first became popular in the last century primarily for building ships and air planes. Driven by economization, welding robots became prevalent in the 1980s and are nowadays ubiquitous in metal processing. Despite the broad adoption of robotic welding, automated quality assessment of the resulting welds is still in its infancy, and therefore, quality checks are commonly performed by human experts. In particular the automotive industry is a driving force in this sector, yet the quality inspection is usually done manually by a human operator. At this point, automatic quality assessment comes into play, which is typically done offline [30, 39] to avoid welding-related image impairments. A review of radiography-based offline quality inspection techniques is given in Silva and Mery [47, 48]. Ultrasonic testing of weld beads and spots is suggested in [9, 33] and a multi-sensor approach to quality monitoring is described in [28]. All these technologies are per definition much more expensive, both in acquisition and maintenance, than the online visual quality inspection presented in this section.

A *welding process* or *task* refers to an automated and repeatable weld. Visual inspection can be performed in an *offline* or *online* manner. An offline inspection denotes an ex situ analysis after completing the weld process. In terms of online analysis, the currently processed welding area or so-called *weld pool* is commonly exploited to identify defects formed during a welding process [51] and to track the seam ahead of the welding torch [3, 15]. A major impediment to online processing is the harsh *image acquisition* environment, characterized by arc-glare, fume, sparks, welding spatter, reflections, etc. To overcome this drawback, additional hardware assisting the camera equipment is indispensable. For example, laser light is used to facilitate the seam tracking ahead of the welding torch in [27, 65].

Another approach is based on recordings taken directly behind the welding head. For this purpose, a camera is positioned behind the welding head providing images of the weld pool and parts of the solidified bead, which is tracked in successive frames. Hence, the term *weld bead tracking* refers to recovering the solidified joint in consecutive frames. In our weld inspection system, similarly to [46], an additional stroboscopic light source assists in suppressing the adverse welding effects on the camera recordings.

The presented online quality assessment framework can be subdivided into several tasks: (i) acquiring suitable images of a welding process, (ii) detection/tracking of the weld bead in each frame, (iii) segmentation of the weld bead, (iv) feature extraction, and (v) matching. The approach presented here is mainly based on the work done by Fronthaler et al. [14]. Following, each aforementioned step is described in more detail and a short summary of *image acquisition* and *weld bead tracking* techniques are presented. A more detailed presentation of our system can be found in Heber et al. [19]. Weld seam *segmentation* in offline scenarios is usually achieved by thresholding [53] or by a global Hough Transform [23] on a binary image containing a straight weld [60]. A *weld bead segmentation* algorithm is also explained, which links the outcome of several local Hough Transforms with two spline functions in each frame. The online requirement of the suggested quality assessment poses greater challenges for the segmentation step. The *quality assessment* is based on *shape features*, which can easily be extracted from the segmented weld bead. Although *texture features* [53, 57] have been suggested in offline analysis of welds, simple *shape features* [46] based on the positional information exhibit sufficient information to assess local anomalies in the weld quality. In offline defect detection, the *classification* task is typically handled by machine learning algorithms such as support vector machines (SVMs) [57, 67], neural networks [1, 56], or other statistical classification algorithms [48]. Considerable drawbacks of these techniques are their complexity and the need for extensive amounts of training data. However, quality decisions based on distance measures between the current weld and a reference reduce the classification to a simpler *matching* task. Furthermore, the local variance of the extracted shape features in consecutive frames can be used as a training phase to learn the variability of the targeted welding process. For enabling a comparison between the currently observed weld bead and a reference, several distance measures, such as dynamic time warping for deformable matching, can be also employed.

There are numerous advantages emerging from the presented online weld quality inspection setup:

- The throughput rate is much higher, if the quality of a workpiece is checked online, because there is no need for a second offline run of the welding robot, or another type of offline check. In particular in the welding industry, efficiency plays a pivotal role.
- The presented methods act exclusively on the newly welded bead behind the welding torch, indicate problems in the welding process, and trigger according to actions (e.g., stopping the welding robot in case of an error). Thus, the

suggested algorithms yield complementary information for real-time welding control, which cannot be gathered by the aforementioned seam tracking only.

- Online monitoring provides additional hints concerning the quality of a welding process, which are not present in case of offline quality checks (e.g., the brightness of the scene). Also an increased amount of spatter and smoke are usually an indication for erroneous welding.
- Finally, in this online scenario, parts of the weld pool can be analyzed, either automatically or by an operator.

3.3.1 Image Acquisition and Weld Bead Tracking

The image acquisition device can be mounted next to the welding torch as shown in Fig. 3.4. It consists basically of two parts: a digital camera and a stroboscope. The former provides 10 grayscale images per second at a given resolution (e.g., 752×480 pixels). To counteract welding-related image impairments (e.g., arc-glare, fume, and welding spatter), a stroboscopic light source acts as a high-performance flash light, which is triggered concurrently with the camera (for example, every 100 ms). Examples of resulting imagery are given in Fig. 3.5. The left image depicts two ordinary steel plates joined by a weld. In the right image, galvanized steel is welded which results in reflections and an increased formation of smoke. In this case, reliable quality assessment is challenging since the view onto the weld is clouded.

The camera is also linked to an industrial PC, where all further processing is carried out. In such case, only the weld bead and its immediate neighborhood serve as features for the employed tracking algorithm. Thereby, the effect of surrounding objects on the tracking parameters can be eliminated. A strongly simplified summary of the algorithm is given in Algorithm 1.

Fig. 3.4 The "Q-Eye" image acquisition device mounted next to the welding torch on a robotic arm. Image courtesy of Fronius Ltd
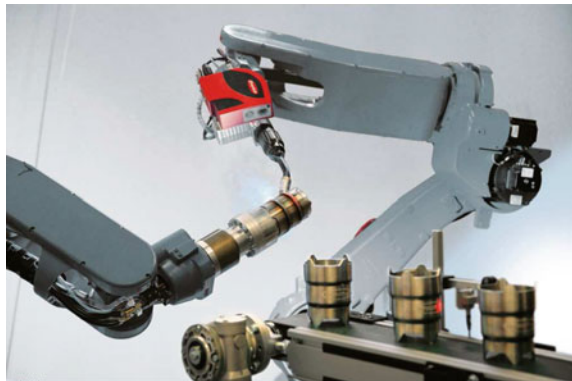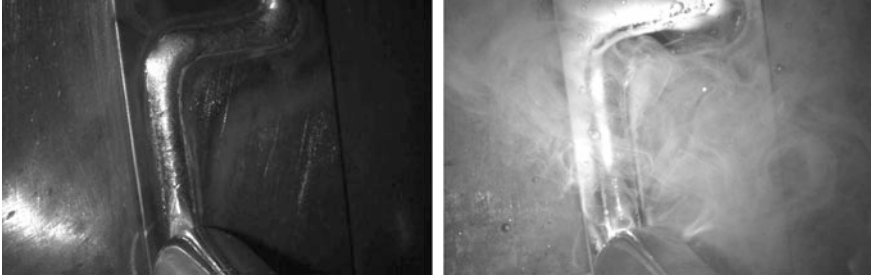
Fig. 3.5 The effect of the surface finish on the image acquisition; welding of galvanized steel (*Right*) results in inferior image quality as compared to ordinary steel (*Left*)

Algorithm 1 Weld Bead Tracking

Require: m initial tracking points TP_0
Ensure: Trajectory TP_k
   1.- Generate a Cubic Spline *CS* by using TP_0
   2.- Move to the subsequent frame $t = t + 1$
   3.- Search for TP_{t-1} in frame t and create an updated CS.
   4.- *Predict* TP_t by evaluating the updated CS at the positions of TP_0 on the initial spline.
   5.- *Correct* TP_t to avoid drifting away from the weld bead.
   6.- Goto step 2 until the welding process ends.

Although the weld bead tracking algorithm is real-time capable, it is only executed once per welding task, because it is assumed that the trajectory of the welding robot remains unchanged—both in position and pace. The resulting tracking information is recorded and replayed whenever the same welding process recurs. Such a strategy ensures an efficient use of the computational resources. Note that it is entirely up to the segmentation and feature extraction stages to collect evidence for welding defects.

3.3.2 Weld Bead Segmentation

The steps involved in the proposed strategy for weld bead segmentation are arranged as illustrated in Fig. 3.6, each of them to be detailed subsequently.

For a set of T image frames of a welding process grabbed over time, the algorithm summarized above provides M tracking points $TP_{t,m} = \{P_{t,m} | t = 0, \ldots, T - 1; m = 0, \ldots, M - 1\}$, where $P_{t,m} = [x_{t,m}, y_{t,m}]$. For segmentation purposes, it is only need a subset $WP_{t,n} = \{P_{t,n} | t = 0, \ldots, T - 1; n = 0, \ldots, N - 1\} \subset TP_{t,m}$, holding those N points per frame, which are situated on the solidified weld. In addition, a set of angles $WA_{t,n} = \{\alpha_{t,n} | t = 0, \ldots, T - 1; n = 0, \ldots, N - 1\}$ describing the

Fig. 3.6 System architecture of the suggested weld bead segmentation
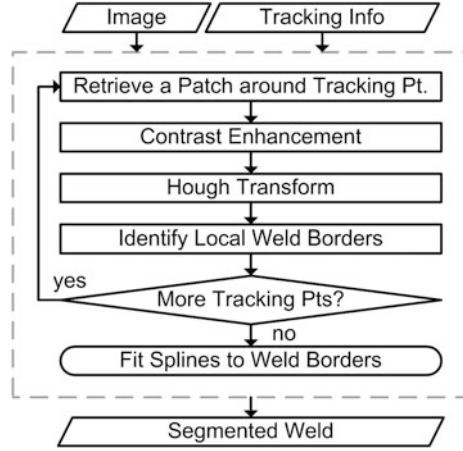
approximate orientation of the weld bead at $WP_{t,n}$ is provided. To facilitate local segmentation, $WA_{t,n}$ is used to extract N rotation normalized image patches $g_{t,n}$ about $WP_{t,n}$ from the current frame t. The right image (image VI) in Fig. 3.7 shows a patch extracted from the left image (image I). Contrary to this example, the weld bead within the image patch is typically slightly skewed and/or displaced from the center. To counteract the demanding illumination conditions and to enable robust feature extraction, the set $g_{t,n}$ is subjected to a local contrast enhancement [6]

$$f_{t,n}(x,y) = \frac{g_{t,n}(x,y) - m_{g_{t,n}}(x,y)}{\sigma_{g_{t,n}}(x,y)} . \tag{3.3}$$

In Eq. (3.3), an image patch $g_{t,n}$ is normalized by locally enforcing a zero mean value and a unit standard deviation. The result of such a local contrast enhancement $f_{t,n}$ is displayed in Fig. 3.7 (image III). Although some random structure emerges from the background, the reflections in the middle of the weld bead are damped and the edges appear more clearly.

The contrast-enhanced image patches $f_{t,n}$, each covering a small and approximately linear tranche of the weld bead in frame t, form the basis of the proposed weld bead segmentation. For further analysis of $f_{t,n}$, N local Hough Transforms are computed to identify the two weld bead borders within each image patch. Based on the work of Gormat et al. [36], each $f_{t,n}$ for edges is searched to preselect candidate lines and to speed up the transform

$$\left| \nabla f_{t,n}(x,y) \right| = \sqrt{\partial_x f_{t,n}(x,y)^2 + \partial_y f_{t,n}(x,y)^2}, \tag{3.4}$$

where $\partial_x f_{t,n}$ and $\partial_y f_{t,n}$ denote the partial derivatives in x and y direction, respectively. The latter are implemented by means of separable Gaussians and their

derivatives. Cartesian coordinates can be transformed into the polar Hough space (aka parameter space, accumulator space, etc.) by

$$\theta_{t,n} = \arctan\left[\frac{\partial_y f_{t,n}(x,y)}{\partial_x f_{t,n}(x,y)}\right], \tag{3.5a}$$

$$\rho_{t,n} = x \cdot \cos\theta_{t,n} + y \cdot \sin\theta_{t,n}, \tag{3.5b}$$

where $\rho_{t,n}$ is the length of a normal vector from the origin to a candidate line in the Cartesian space and $\theta_{t,n}$ corresponds to the orientation of this vector with respect to the x-axis. Each line in the (x,y)-space is characterized by a point in (ρ,θ)-space and therefore collinear Cartesian points result in an accumulation about the corresponding (ρ,θ) parameters in Hough space. For this purpose, the edge strength $|\nabla f_{t,n}(x,y)|$ is exploited as aggregation bias, favoring reliable edge information. Furthermore, the threshold condition $|\nabla f_{t,n}(x,y)| > \tau_e$ has to be fulfilled at a certain pixel location (x,y) in Eqs. (3.5a) and (3.5b) to be accounted for in the transformation. Image IV in Fig. 3.7 depicts the Hough space of the contrast-enhanced image patch to the left. One can clearly spot two local maxima, which correspond to the borders of the weld bead tranche.

Unfortunately, lines within $f_{t,n}$ are far from ideal, which obviously results in noisy representations in the Hough space. Nevertheless, some simplifications due to prior knowledge can be made to enable the identification of the local weld bead borders within the (ρ,θ)-space of each $f_{t,n}$: (i) Only one weld bead is expected, (ii) restrictions concerning the position and skew, and (iii) exactly two lines at a predefined distance range (width) are present.

The reverse transformation (aka de-Houghing) of the two local accumulator maxima into Cartesian coordinates yields two lines approximating the weld bead borders. As indicated in image V of Fig. 3.7, one point per line is of special interest, i.e., the intersection with a normal passing through the center of $f_{t,n}$ which corresponds to $WP_{t,n}$. Thus, two additional point sets $B1_{t,n}$ and $B2_{t,n}$ are obtained,
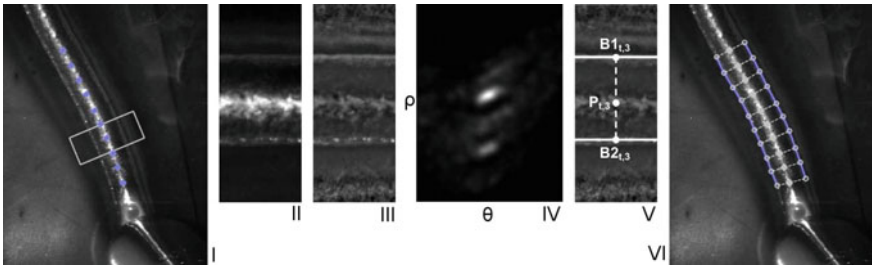
Fig. 3.7 Weld bead segmentation visualized by an example frame at time D; *I* 10 tracking points $WP_{t,0-9}$ on the weld bead, *II* example of a rotation normalized image patch *highlighted* to the left ($g_{t,3}$), *III* contrast-enhanced version of the image patch ($f_{t,3}$), *IV* Hough space representation, *V* weld borders (corresponding to the two peaks in hough space) superimposed on $f_{t,3}$, *VI* the segmented weld bead

sampling the shape of the two weld bead borders. Note that all lines can be reconstructed by means of $B1_{t,n}$, $B2_{t,n}$, and $WP_{t,n}$ only.

In the final step of the suggested weld bead segmentation algorithm, the *cubic parametric spline curves* $S1_t$ and $S2_t$ are built to interpolate the control point sets $B1_{t,n}$ and $B2_{t,n}$, respectively, for a frame at time t. This yields a continuous segmentation of the weld bead from the background. The spline curve $S1_t$ is defined piecewise over $N-1$ subintervals $b1_{t,n} = [B1_{t,n}, B1_{t,n+1}]$ by the two polynomials $\left(X1_{t,n}, Y1_{t,n}\right) : b1_{t,n} \to \mathbb{R}^2$, where $n = 0, \ldots, N-2$. The so-called cord length is used to parameterize $S1_t$, which approximates the length of the spline curve by linearly connecting adjacent points in $B1_{t,n}$. By using a two-dimensional description of $S1_t$ via $\left(X1_{t,n}, Y1_{t,n}\right)$, it is possible to model arbitrary curves formed by a weld bead at frame t. The chosen spline type is characterized by polynomials of a degree up to 3, joining with continuity C^2. Clearly, $S2_t$ is defined in the same way. A derivation of the actual polynomials $X1_{t,n}$, $Y1_{t,n}$, $X2_{t,n}$, and $Y2_{t,n}$ as well as the description of an appropriate computation algorithm is beyond the scope of this chapter and thus the interested reader is referred to [12] for a detailed disquisition on splines. The two splines, each based on $N = 10$ control points, are depicted in the last image of the segmentation example given in Fig. 3.7 (image VI).

3.3.3 Quality Assessment Features and Measures

The quality of a weld can be defined in terms of its strength in comparison with the material it joins. However, the exact quality and appearance requirements rest upon the application at hand. In offline scenarios, the quality of a specimen can be tested either by *nondestructive* techniques (e.g., based on radiography or ultrasound) or in a *destructive* manner to determine the ultimate stress limits [59]. Based on the online vision system described in Sect. 3.1, the suggested quality inspection can be assigned to nondestructive testing. In the following, we briefly present the employed shape features and elaborate on several reduced-reference quality measures, some of which require training.

3.3.3.1 Shape Features

Having the segmented weld bead within each image frame at one's disposal, the actual quality assessment can be carried out. To do that, a feature comparison is needed. When a real-time requirement is present, *shape-based features* are preferred over *texture features* which are more complex. Besides, the former provides higher stability than the latter. Due to the demanding scene lighting, the weld texture can change drastically even from one frame to the next. By using shape-based features, it is possible to reuse the sampled boundaries $B1_{t,n}$ and $B2_{t,n}$ for a local shape description in each frame t. The *weld bead width*, computed by
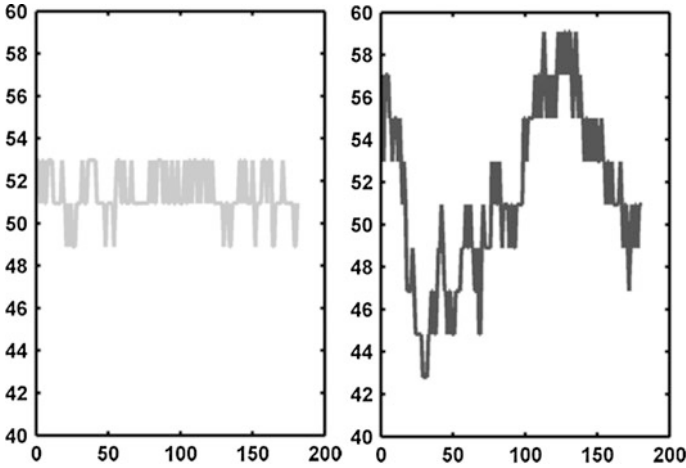
Fig. 3.8 Bead width measurements of two welds (same welding task); (I) error-free sample, (II) defective sample due to welding current fluctuations. The x-axis corresponds to the frame number and the y-axis shows the width in pixels

$$w_{t,n} = ||B1_{t,n} - B2_{t,n}||, \tag{3.6}$$

is used as input for the suggested quality measures. Considering one point on the physical weld, it recurs several times at varying pixel locations in different frames. Due to this natural redundancy, already subsets of $B1_{t,n}$ and $B2_{t,n}$ suffice to capture the local bead width reliably. However, if the quality decisions were based on a single image instead of a continuous video feed, such simplifications could not be made. In that case, it is advisable to exploit the complete shape information encoded in the spline trajectories.

Figure 3.8 shows bead width measurements of two welds, both originating from the same welding task. The graph to the left depicts the error-free case, where only small variations can be observed. A defective sample exhibiting much stronger fluctuations is displayed to the right. The disturbance is caused by a reduction followed by an increase of the welding current, which influences the bead width noticeably. The simple shape features $w_{t,n}$ described above enable the detection of welds of inconsistent width and crooked beads. Another indication for defective welding is the occurrence of excessive fume or spatter. Due to the online setup, it is possible to observe these abnormalities in terms of irregular feature values.

3.3.3.2 Non-reference-Based Quality Assessment Measures

In real-world applications, it is not always feasible or practical to provide some sort of reference data. If a weld bead is laid only once or just a few times, it might not be worth the effort to produce ground-truth data. However, the need for automatic

quality inspection persists. For this purpose, two simple ad hoc quality assessment measures are suggested, not demanding any reference information. Both exploit the fact that the width of high-quality welds varies smoothly over time. On the other hand, the weld borders in erroneous regions are not well-defined. Therefore, the width changes rapidly, resulting in large distance measures.

Absolute Difference

The first quality measure $d1$ is based on the *absolute difference* of the feature values from the current frame to the corresponding values in the previous frame

$$d1_{t,n} = \left| w_{t,n} - w_{t-1,n} \right|. \tag{3.7}$$

Thus, a value $d1_{t,n}$ is large, if the bead width changes rapidly in consecutive frames. On the other hand, smooth width changes over time results in small values of $d1$.

Variance

The second ad hoc quality measure $d2$ computes the *variance* of the feature values over the s preceding time frames

$$d2_{t,n} = \frac{1}{s+1} \sum_{k=t-s}^{t} \left(w_{k,n} - \overline{w}_{t,n} \right)^2, \tag{3.8}$$

where $\overline{w}_{t,n}$ denotes either the mean value of $w_{t,n}$ across a time window

$$\overline{w}_{t,n} = \frac{1}{s+1} \sum_{k=t-s}^{t} w_{k,n} \tag{3.9}$$

or a predefined value in pixels (e.g., 80). Similarly to $d1$, it is expected the local variance $d2$ to be low in case of good quality. If $d2_{t,n}$ exceeds a predefined threshold, it is assumed that an error occurred within the corresponding time window during the welding process.

3.3.3.3 Reference-Based Quality Assessment Measures

The desired weld widths at several predefined positions per frame yield a resilient shape model that enables accurate quality inspection. To match the candidate information with the given reference, two quality measures are suggested. The first quality measure is a direct comparison based on the Euclidean distance and the second one incorporates dynamic time warping.

Euclidean Distance

The local Euclidean distance $d3$ between a test sequence $w_{t,n}$ and a reference $r_{t,n}$ is the square root of the sum of squared differences within a local window of size s

$$d3_{t,n} = \sqrt{\sum_{k=t-s}^{t} (w_{k,n} - r_{k,n})^2} \,. \tag{3.10}$$

The sequence w should be as similar as possible to r, which results in small values of $d3$. If w and r are not of equal length, one of them needs to be shortened.

Dynamic Time Warping

The *dynamic time warp* distance (DTW) [43], in contrast to the Euclidean distance measure, allows for a comparison between time series under a nonlinearly warped time axis. For this reason, DTW is popular in many application areas [32, 40]. Signal delays or slight periodic differences do not influence the distance between the time series. In our case, this translates to tackling slight asynchronism as well as local speed deviations between the candidate and reference information caused by the welding robot.

The aim of dynamic time warping was to find the optimal matching elements of two time series w_t and r_t so that the total sum of Euclidean distances between these elements is minimized

$$m_t = \arg\min_{\vec{i},\vec{j}} \sum \text{dist}(w_i, r_j) \,, \tag{3.11}$$

where $m_t = [i_t, j_t]$. To find this set of matches, three steps need to be taken. In the first step, the distance matrix dm between all elements of the two time series is calculated by

$$dm(i,j) = |w_i - r_j| \,. \tag{3.12}$$

Assuming the matching vector m_t forms a continuous path through dm; the problem of finding this vector can be solved using the dynamic programming principle in polynomial time. To find this optimal continuous path, the minimum cumulative distance matrix D is computed in the second step. The initial values of D are set to dm and starting at position $(1,1)$, $D(i,j)$ is computed as

$$D(i,j) = dm(i,j) + \min[D(i-1,j), D(i,j-1), D(i-1,j-1)] \,. \tag{3.13}$$

The value $D(i,j)$ represents the minimum sum of distances to reach the point (i,j) in the matrix. The value at the lower right corner of D is the minimum distance sum as determined by the optimal matching vector. In the third step, the process is reversed to find the actual matching vector m_t. Starting at the lower right element of

D, the minimum path taken through the matrix is traced back to the upper left element. This is done by selecting and storing the coordinates m_t of the minimum cumulative distance matrix elements.

The DTW distance resulting from this procedure is a global measure of distance between the sequences. Using the matching vector m_t, a local DTW distance measure $d4_{t,n}$ can be defined as follows:

$$d4_{t,n} = D_n(i_t, j_t) - D_n(i_{t-s}, j_{t-s}) \ . \tag{3.14}$$

Thus, the quality measure $d4_{t,n}$ computes the local DTW distances of a sequence w from a given reference r, which corresponds to the sum of the matched sample distances over a certain time window s.

3.3.3.4 Summary

In this section, we proposed a reduced-reference online weld quality assessment framework and demonstrated its effectiveness for automated welding processes. The online image acquisition is enabled by the stroboscopic light source, which assists in suppressing unfavorable welding effects such as arc-glare, spatter. The employed tracking algorithm yields reliably located points on the weld bead. Although the Hough Transform is not designed for seeking vague lines, we can identify the borders even in case of curved welds. This is due to the use of edge information coupled with our local concept, where only short tranches of the weld are inspected at a time. The two spline curves yield a smooth interpolation of the sampled borders and thereby enable a proper segmentation of the weld in each frame. Already very simple shape features based on the local weld bead width hold sufficient information to reliably separate error-free samples from the defective ones in a large database. Despite achieving slightly higher error rates, the non-reference-based quality assessment measures are highly valuable since they afford ad hoc inspection wherever manufacturing reference data is not an option.

3.4 Rail Surface Analysis

In order to avoid severe accidents, a frequent and adequate inspection of rails is a necessary precaution. Traditionally, rail inspection was done by visual means, e.g., in combination with hammer sounding, and later automated inspection methods based on, e.g., ultrasonic or magnetic sensors, and X-ray were used. Although some problems are not detected by visual means, automated visual inspection offers additional information and a denser coverage when compared to established methods. In particular, the analysis of rail wear resulting from rolling contact fatigue (RCF) [13] is well suited to high-resolution visual inspection of the rail head. So-called spallings or headchecks result from small break outs on the rail
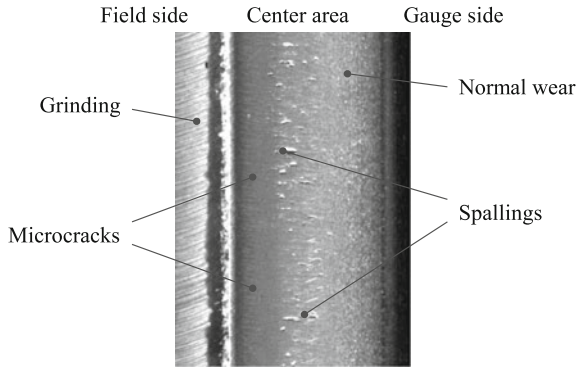
Fig. 3.9 Surface regions observed from a *top-down* view of a rail head surface and manual classification: The gauge side is the inner side of the rail in contact with the wheel, microcracks and spallings typically occur in the center area, the field side is the non-contact area showing grinding artifacts in this case

surface and are often related to RCF [66]. Figure 3.9 shows regions and typical surface characteristics as observed by a top-down view of a rail head.

In this section, we investigate detection of small surface disruptions and mapping of the rail surface into regions by classification.

3.4.1 Rail Crack Detection

Spallings or headchecks are commonly regarded to be preliminary stages of cracks and even rail breaks. Furthermore, preliminary stages of spallings and headchecks are the so-called microcracks. Microcracks originate from elastic flow and wear away of the rail and show a characteristic appearance. In the studied data, microcracks become visible for resolutions better than 0.25 mm per pixel. Grinding is a measure to restore rail profile and remove distortions in order to avoid further damage. The application of grinding wheels results in a very specific surface texture. When the rail is run again after grinding the grinding texture disappears in the center and gauge regions, and remains visible only at the field side of the rail.

In 2.5D analysis, the red and the blue channel of a color line-scan camera image are used. Illuminating line light sources is placed parallel to the line-scan camera. The lights and the camera itself are mounted on a railway vehicle and oriented orthogonal to the rail direction. Figure 3.10 shows the rail car with mounted inspection system and a detailed view of an acquisition box. The corresponding analysis scheme is shown in Fig. 3.11a. The setup is based on a dark-field illumination principle, i.e., for undistorted surface patches only a small amount of light is reflected toward the sensor, whereas surface disruptions typically cause reflections.
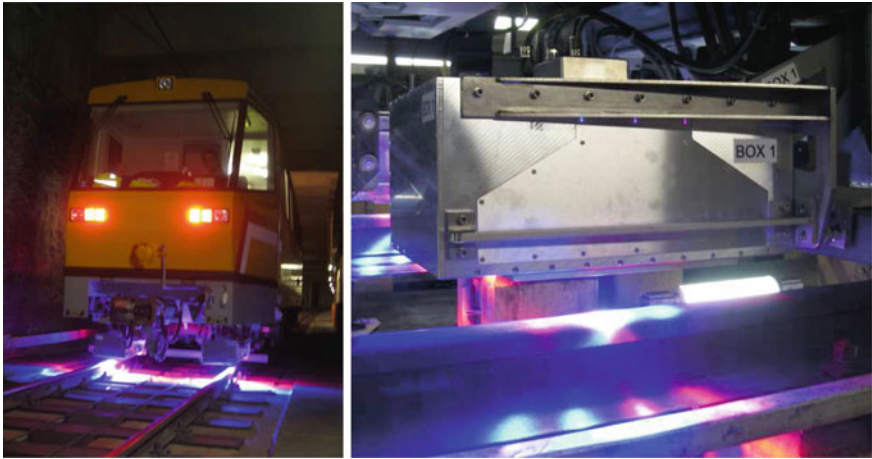
Fig. 3.10 Rail crack inspection system: system mounted on subway car (*Left*), detailed view of acquisition device (*Right*)
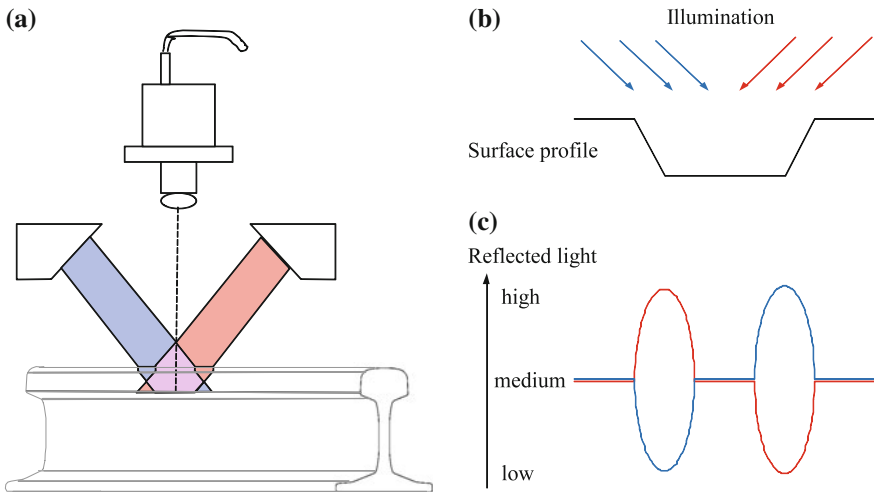
Fig. 3.11 Acquisition setup and model of reflection properties: **a** *top-down* view of the head surface using a line camera and illumination by different line light sources under oblique angles, **b** distorted surface profile into driving direction and direction of sources of illumination, **c** model of reflectance for a distorted surface profile

One light source, e.g., a red one, is placed in front and another light source, e.g., a blue one, is placed behind the line camera with respect to the driving direction. The angle of specular reflection from undistorted rail surfaces into the camera is avoided, and strong reflections and shadows are observed in cases of surface disruptions only. The expected reflection properties due to a surface

distortion, i.e., a spalling, are shown in Fig. 3.11b. Light from two spectral channels, e.g., from a blue light source at the left and a red light source at the right, approaches the object surface at oblique angles. The shown surface profile is taken along the rail direction. Due to the dominating specular reflection on metal surfaces, light in the red channel is typically strongly reflected back to the camera when hitting the left edge of the surface disruption, see Fig. 3.11c. On the other hand, no specular reflection happens in the red channel on the edge on the right. The blue channel shows an opposite behavior when compared to the red channel, and this property is termed anti-correlation. More details of this approach can be found in Huber-Mörk et al. [24].

3.4.1.1   Anti-correlation-Based Analysis

The identification of surface disruptions is formulated similar to a target detection problem, i.e., a target T is identified if it is sufficiently different from its background B statistics. In the context of target detection theory, this is expressed by the ratio of maximum a posteriori probabilities (MAPs) $P(T|s)/P(B|s) > 1$ for target and background, with s being the observed feature. Application of Bayes rule and Neyman-Pearson theory results in the likelihood test $P(s|T)/P(s|B) > t$, where t is a threshold [44].

In our case, a simple procedure involving four target detectors of the form $s/\bar{s} > t_s$ and $s/\bar{s} < 1/t_s$, where s is a measurement corresponding to a potential target and $\bar{s}$ is the estimated background, is proposed. Referring to Fig. 3.11c, we define the *detection position* to be in the middle of the surface profile. With $s = \{r, b\}$, where r and b are red and blue light received by the camera, the four target detection equations become

$$r_{+d}/\bar{r} < 1/t_r, \quad r_{-d}/\bar{r} > t_r,$$
$$b_{-d}/\bar{b} < 1/t_b, \quad b_{+d}/\bar{b} > t_b. \tag{3.15}$$

The subscript $-d$ indicates a position in driving direction, e.g., at a distance of d pixels, with respect to the detection position. The position $+d$ is located d pixels looking back in driving direction w.r.t. the detection position. The r_i, b_i are the red and blue values for the pixel at position i and $\bar{r}, \bar{b}$ are backgrounds estimated by local smoothing. Two different thresholds t_r, t_b, related to the different color channels, are defined. The combination of the four target detectors is done by taking into account the anti-correlation property, i.e., the presence of a target is assumed only if all four detectors coincide in their decision for a target. In order to stabilize detection for larger surface disruption, we integrate the detectors over scales, e.g., we allow a range of displacements $d = d_{min}, \ldots, d_{max}$ for which we test the detection equations from Eq. (3.15).

3.4.1.2   Cross-Channel Co-occurrence-Based Analysis

In this improved approach, we integrated detection results over meaningful regions, so that anti-correlation regions with a mutual relation to each other are treated together. Similarly to the target detection principle introduced in the previous section, we used properties related to deviation from the local image background.

According to the model, at least some parts of a deviation due to a surface disruption exhibit anti-correlation configurations. These could be a set of distributed small spots, which are prone to be eliminated as irrelevant when considered as single detections. Using the information of background deviation clings all occurrences of anti-correlation configurations regarding the same disruption together, no matter if these configuration regions are adjacent. In this manner, a background deviation object is considered a rail surface disruption, if a significant amount of member pixels of that deviation region displays the anti-correlation model.

Background deviation is derived from the local image background computed as the convolution of the original image I with a large averaging filter h, i.e., $I_B = I \times h$. In the used data, a 21×21 box filter showed satisfying results. The background image I_B is then subtracted from the original image I giving the deviation image ($I_D = I - I_B$). The resulting difference image I_D is comprised of the background deviations. By means of a connected-component extraction algorithm, applied to the luminance image of I_D, we extracted the deviation regions and evaluated the amount of anti-correlation configurations within their region boundaries in I_D.

As a measure of the score of anti-correlation configurations in a region of interest, we determine the relative frequencies of pixel pairs within the region, which fulfill the anti-correlation condition, i.e.,

1. both pixels are within the region of interest,
2. both pixels are in the same column but in different lines,
3. the upper pixel is significantly more blue than red,
4. the lower pixel, at the same time, is significantly more red than blue.

That means, we are interested in counting vertical pixel pair configurations showing the co-occurrence of an upper blue and a lower red pixel partner. Recall that the anti-correlation property appears along rail driving direction, which means vertically in the image columns in consequence of the setup of the light sources and the line-scan camera.

A classical method for representing the co-occurrence of special brightness configurations of pixel pairs in image patches is the co-occurrence matrix [18]. Usually, it is used for texture description in grayscale images. For a special grayscale image patch G and a fixed pixel distance d, a grayscale co-occurrence matrix [49] may be expressed in a non-normalized form as follows:

$$P_d(v_1, v_2) = |\{((k,l),(m,n)) \in G \times G :$$
$$m - k = d, \; l = n, \qquad\qquad\qquad (3.16)$$
$$I(k,l) \in I(v_1), \; I(m,n) \in I(v_2)\}|,$$

where $|\{\ldots\}|$ refers to set cardinality and $I(v_1)$, $I(v_2)$ are brightness value intervals (i.e., quantization intervals). In the original formulation, co-occurrence matrices also depend on an angle ϕ, which indicates in which direction d is meant. In our application, only the vertical direction is of interest, why we omit the use of ϕ in our descriptions.

However, $P_d(v_1, v_2)$ describes how frequently two pixels with gray levels in $I(v_1)$, $I(v_2)$ appear in the region G. In order to fit our problem at hand, we propose to use the concept of co-occurrence matrices to describe the frequencies of anti-correlation configurations of the extracted regions of interest (G) for characterization of rail surface disruptions. Thus, we introduce cross-channel co-occurrence matrices for color images (see 3.17), in which co-occurrences in two different color channels, particularly differences of color channels ($b - r$ and $r - b$), are considered, rather than only in one grayscale image:

$$C_d(v_1, v_2) = |((k,l),(m,n)) \in G \times G :$$
$$m - k = d, \; l = n, \qquad\qquad\qquad (3.17)$$
$$I_D^{b-r}(k,l) \in I_D^{b-r}(v_1), \; I_D^{r-b}(m,n) \in I_D^{r-b}(v_2)\}|,$$

where I_D^{b-r} is the difference of the blue color channel of I_D minus its red color channel, and I_D^{r-b} is the $r - b$ color channel difference of I_D.

The presented definitions of co-occurrence matrices so far describe absolute frequencies of pixel configurations. Yet, deviation regions are of different sizes. To be able to compare co-occurrence matrices for regions of different sizes, we need to normalize $C_d(v_1, v_2)$ by its matrix sum (i.e., the total amount of pixel configuration pairs considered), giving a cross-channel co-occurrence matrix of relative frequencies:

$$\bar{C}_d(v_1, v_2) = \frac{C_d(v_1, v_2)}{\sum_{i,j} C_d(i,j)} \; . \qquad\qquad (3.18)$$

$\bar{C}_d(v_1, v_2)$ reflects the relative frequencies of pixel configurations in G for one fixed pixel distance d. As mentioned in the previous section, rail surface disruptions can comprise anti-correlation configurations for a variety of distances d. Therefore, for each region of interest, we compute a set of matrices $\bar{C}_d(v_1, v_2)$, where $d = d_{\min}, \ldots, d_{\max}$ and $d_{\min}, d_{\max}$ are minimum and maximum vertical distance of pixel configurations in that region of interest. Then, we compute the average over all these matrices, in which way we take account for any possible anti-correlation configuration distance d:

$$\bar{C}(v_1, v_2) = \frac{1}{d_{max} - d_{min} + 1} \sum_{d=d_{min}}^{d_{max}} \bar{C}_d(v_1, v_2). \tag{3.19}$$

With the cross-channel co-occurrence matrix $\bar{C}$, we have generated a tool for determining the relative frequencies of anti-correlation configurations that are present in a region G of I_D. As anti-correlation is the co-occurrence of high $b - r$ values and high $r - b$ values in distance d, one would expect to measure a significant amount of entries apart from the first few lines and first few columns only for disruption regions. Consequently, the criterion to characterize disruptions is based on the sum of entries in a right, lower sub-matrix of $\bar{C}$:

$$G \text{ is surface disruption if } \left(\sum_{i=i_0}^{n} \sum_{j=j_0}^{n} \bar{C}(i,j) \right) > t , \tag{3.20}$$

where $i_0 > 1$, $j_0 > 1$ (usually $i_0 = j_0$), and $i = 1$, $j = 1$ indicate the first line and column of $\bar{C}$, respectively; t is a threshold value that has been determined in experiments.

Detection results for rail surface disruption are shown in Fig. 3.12. Green rectangles indicate background deviations comprising a significant amount of anti-correlation configurations (i.e., surface disruptions), whereas red rectangles indicate general background deviations (often a result of texture resulting from uneven textured surface structure). Small unmarked anti-correlation configuration regions were ignored as part of the false alarm elimination procedure, because they were too small and not part of any bigger background deviation. More details of this approach can be found in [50].

Fig. 3.12 Results for rail surface disruption detection: *Green rectangles* indicate background deviations comprising a significant amount of anti-correlation configurations (i.e., surface disruptions), whereas *red rectangles* indicate general background deviations (often a result of texture resulting from uneven textured surface structure)
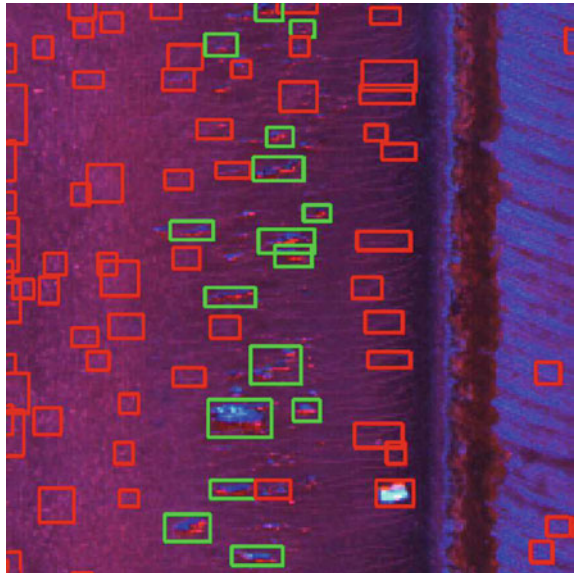
3.4.2 Rail Surface Classification

A number of surface classes describing the condition of rails are summarized in a hierarchical surface class repository (see Fig. 3.13). Classes of major interest are microcracks, grinding marks, headchecks, spallings, and cracks. Grouping of classes was performed by expert assessment; especially all non-critical classes were collected into a single class.

For surface texture description, we use two-dimensional Gabor filters defined by [17]

$$g(x,y) = \exp\left(-\frac{x'^2 + v^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \phi\right) \qquad (3.21)$$

with

$$x' = x\cos\theta + y\sin\theta,$$
$$y' = -x\sin\theta + y\cos\theta. \qquad (3.22)$$

Filter orientation is determined by θ. The parameter $\phi = 0$ results in a symmetric filter mask, whereas $\psi = -\pi/2$ is the asymmetric filter mask. The Gaussian envelope of the filter is specified by σ and the spatial aspect ratio v. Commonly, a
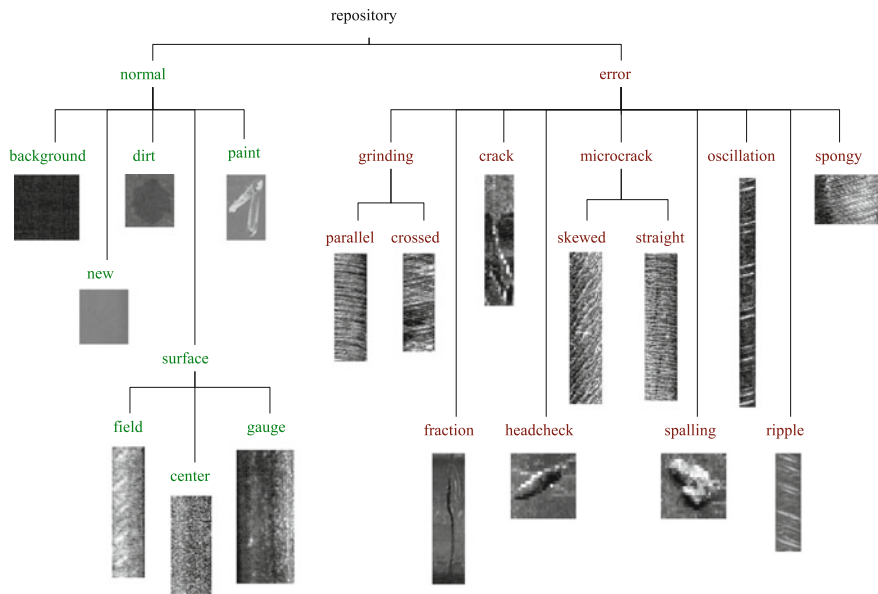
Fig. 3.13 Repository of rail head surface classes: Normal or non-critical surface conditions are shown in the *left* branch; erroneous surface classes are shown in the *right* branch

symmetric and an asymmetric filter mask are allocated into the real and imaginary parts of a complex function, thus forming a quadrature image filter [25].

Invariance to illumination change, in our case mainly induced by the curvature of the rail head, is obtained by image normalization. Pixel-based normalization to zero mean and unit standard deviation with respect to a local window was applied. Similar procedures, e.g., normalization to zero mean and L_1-norm [54], are commonly used.

The selection of orientations and scales of the Gabor filters mainly depends on the application. In order to cover the frequency of the observed undulations of the textures of interest, namely microcrack and grinding texture, on the considered images the scales with wavelength $\lambda = 3, 6, 12$ and orientations of $\theta = 0, \pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6$ were chosen.

The class conditional likelihoods for classes $C \in \{1, \ldots, \Omega\}$ are expressed by a Gaussian mixture model (GMM) [7]

$$p(\mathbf{x}|C) = \sum_{k=1}^{N_C} \alpha_k p_k(\mathbf{x}|C), \tag{3.23}$$

where $p_k(\mathbf{x}|C)$ are Gaussian densities with mean vectors μ_k and full covariance matrices Σ_k. The N_C is the number of components used to model class C. The expectation maximization (EM) [11] algorithm was used to estimate $p(\mathbf{x}|C)$ from data labeled by experts. A Bayes classifier determines class probabilities through

$$P(C|\mathbf{x}) = p(\mathbf{x}|C)P(C), \tag{3.24}$$

where background knowledge might be represented by the class prior $P(C)$.

We introduce background knowledge in a two-step framework. In the first step, we estimate $P(C|\mathbf{x})$ from Eq. (3.24) and condition $P(C)$ on the across-track position y in a second step. Therefore, the Bayes classifier rule becomes as follows:

$$P(C|\mathbf{x}, y) = p(\mathbf{x}|C)P(C|y) . \tag{3.25}$$

The $P(C|y)$ are locally derived, i.e., with respect to along-track direction, from the frequency of class labels $h(C, y)$ assigned in the first step

$$h(C, y) = \sum_{t=t_s}^{t_e} \frac{\delta(L(y, t), C)}{t_e - t_s + 1}, \tag{3.26}$$

where L is the label image $L \in \{1, \ldots, \Omega\}$, and t_s, t_e with $t_e > t_s$ are the along-track coordinates. The expression $\delta(L(y, t), C)$ simply indicates whether class C is assigned for across-track position y and along-track position t. Using Eq. (3.25), the final classification is obtained.

Figure 3.14 shows estimated $P(C|y)$ depending on the across-track position y, e.g., the image column coordinates. For small y, which is the gauge side of the rail,
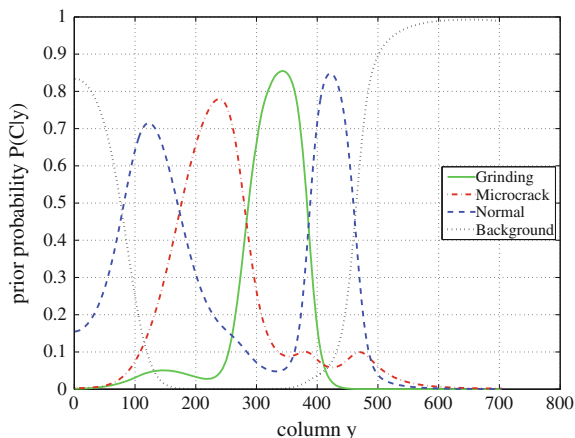
Fig. 3.14 Representation of background knowledge $P(C, y)$ expressed by prior class probabilities for all classes of interest and conditioned on across-track position. The gauge side of the rail head is found on the *left* and field side is located on the *right*

and large y, background is most likely. The across track rail profile is indexed by y. The column index y for the gauge side of the rail ranges from 1 to approximately 200 where normal wear is predominant. The rail head central region column indices ranging from 200 to 400 are dominated by microcracks and grinding. The field side of the rail starts at $y \approx 400$ where normal conditions followed by background are observed.

On the left of Fig. 3.15, a small portion of a rail image obtained in the red channel is shown. The middle image of Fig. 3.15 shows the corresponding classification including fusion with background knowledge. The background knowledge was obtained from the evaluation of a histogram, see Eq. (3.26) taken over 5000 lines. The right image in Fig. 3.15 shows the result with a reject option on the posterior probabilities using Eq. (3.25). The reject threshold was set to 0.5.
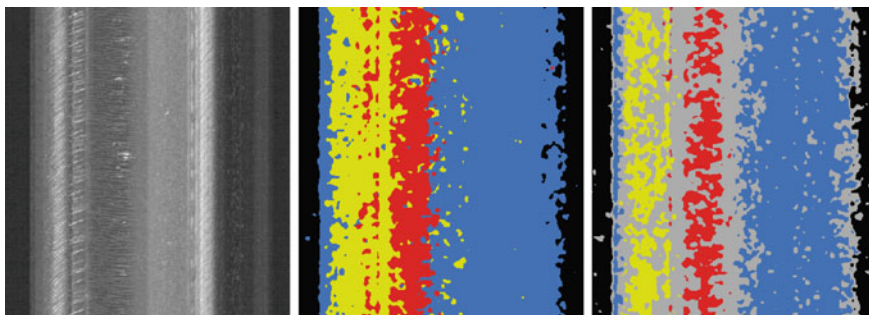
Fig. 3.15 Results of rail head surface classification: Image acquired in *red channel*, classification and classification with rejection Legend: background *black*, normal rail surface *blue*, grinding area *yellow*, microcrack area *red*, no decision (rejection) *gray*

3.5   Multi-view Methods

In order to complement the previously described 2.5D-based analysis, two approaches for 3D analysis especially suited for industrial applications are presented. A multi-view approach suitable for the inspection of metal foils on printed matter, e.g., security features such as optical variable devices (OVD), is introduced. A light-field approach for acquisition of multiple views of the same object position, which enables computational focusing and depth estimation, is applied to coin inspection.

3.5.1   Mirror-Based Multi-view Line-Scan Analysis

We demonstrate the design and setup for a line-scan stereo image acquisition system using a single area-scan sensor, single lens, and two planar mirrors attached to the acquisition device. The acquired object is moving relatively to the acquisition device and is observed under three different angles at the same time. This setup enables surface reconstruction using stereo image processing as well as inspection of object properties varying with the observation angle.

The proposed acquisition device consists of a single area-scan camera, two mirrors, a conveyor belt, on which the inspected object is placed, and two sources of light using line converters as shown in Fig. 3.16. Details of the used hardware components are summarized in Table 3.1. We employed two line converters illuminating the acquired image line from two different angles along the transport
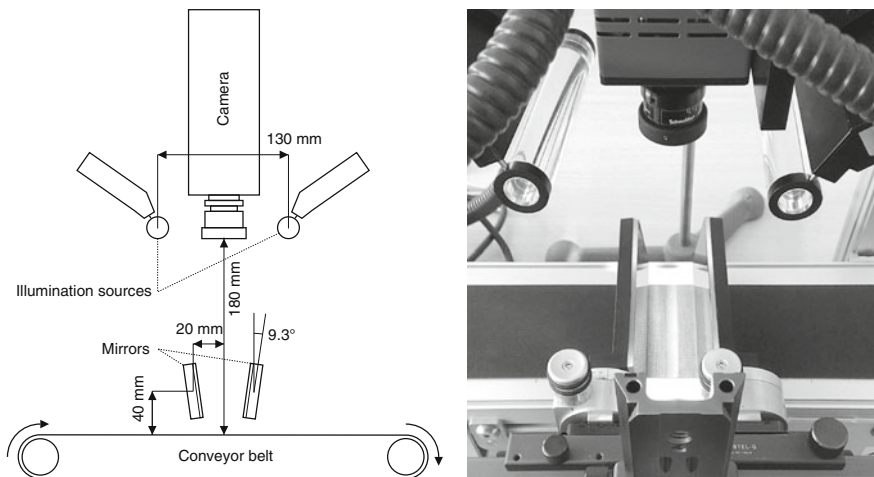
Fig. 3.16 The drawing and photograph of the proposed multi-line-scan acquisition system. Important dimensions of the system are provided for reference

Table 3.1 Parameters of the stereo acquisition setup

Given parameters		
Camera distance	h	180 mm
Lateral view angle	α	8°
Spatial resolution	p	0.083 mm/pixel
Depth resolution	z	0.083 mm/unit disparity
Derived parameters		
Spatial to depth resolution ratio	r	1
Stereo baseline	b	180 mm
Object view angle	β	26.6°
Mirror tilt	γ	9.3°
Mirror shift	d	40 mm
Mirror offset	m	20 mm

direction. The lines are concurrently extracted from the sensor, which is equivalent to placing three single line-scan cameras at different observation angles, see Fig. 3.17 on the left. In order to maximize the acceptable height range for an object under inspection, we decided for an optimized illumination setup, which reflects the light through the mirrors as shown in Fig. 3.17 on the right. More details of this approach can be found in [21].
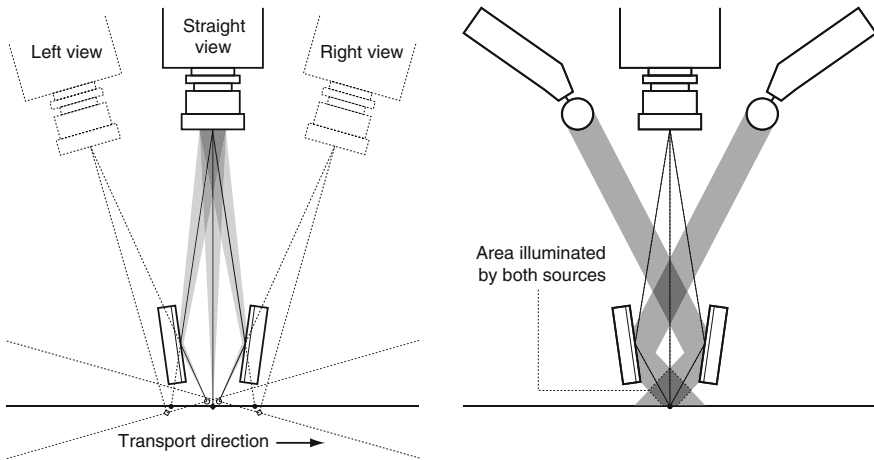
Fig. 3.17 The projection of three single acquired lines to three different sensor lines (equivalent to virtual cameras) by means of mirrors (*Left*). The illumination setup and path using reflection from the mirrors (*Right*)

3.5.1.1   View-Dependent Analysis

OVDs are security features used nowadays widely applied in the banknote printing industry, document security, etc. [42]. OVDs are very robust against counterfeiting, as they change appearance with changing viewing or illumination angles. However, it is the same effect which makes the inspection of these features quite difficult, unless a multi-view acquisition system is employed. As the suggested approach realizes a multi-view system, we found it well suited for the inspection of OVDs. In Fig. 3.18, we provide results obtained by the suggested setup when applied to the hologram feature present at the EUR-50 banknote. One can notice that each of the three views reveals significantly different details of the hologram, which would not be disclosed by any standard acquisition system comprised of a single camera under constant illumination. Utilizing our three views of the hologram, an inspection can be performed with much higher confidence when compared to any standard system.

3.5.1.2   Surface Reconstruction

Apart from the analysis of objects varying with observation direction, the suggested setup is also suited for multi-view stereo reconstruction. In the following, we provide details of the stereo geometry.

In our setup, the mirrors are positioned symmetrically around the optical axis of the camera in such a way that their facing surfaces are orthogonal to the transport direction. The relationships between parameters are expressed by the following set of formulas:

$$
\begin{aligned}
r &= p/z \\
b &= h \cdot r \\
\beta &= \arctan(r/2) \\
\gamma &= (\arctan(r/2) - \alpha)/2 \\
d &= \frac{h \cdot \tan(\alpha)}{r/2 + \tan(\alpha)} \\
m &= d \cdot r/2
\end{aligned}
\tag{3.27}
$$
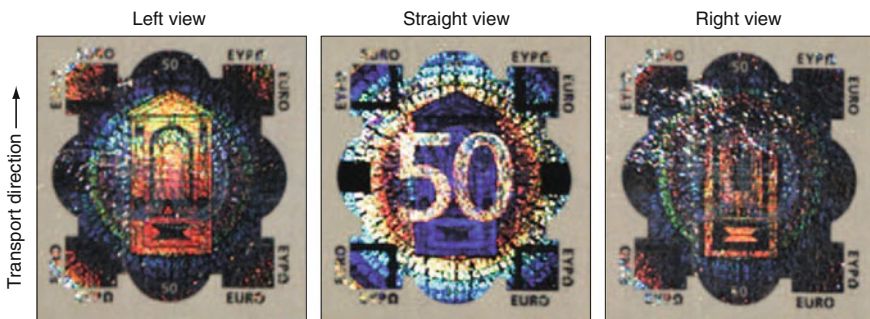
Fig. 3.18 Three views of the hologram at the EUR-50 banknote

where h, α, p, and z are given parameters of the system, see Table 3.1. Note that the spatial and depth resolutions p and z are measured in mm/pixel and mm/unitdisparity, respectively.

With increasing the tilt angle of the mirrors, the depth resolution of the system rises. Theoretically, very high depth resolutions can be achieved by using extreme tilt angles; however, in such cases, the effect of the prolonged optical path in the skewed views becomes indispensable, resulting in severely defocused skewed images. Moreover, stereo matching of wide baseline images becomes more difficult, especially when dense correspondences are required. Another problem could be the gradually decreasing resolution of the skewed images due to the prolonged optical path. It is therefore recommended to keep the tilt angle of the mirrors as low as possible while keeping in mind the desired depth resolution.

The relationship between the tilt angle of the mirrors γ, the lateral view angle α, and the achievable spatial to depth resolution ratio $r = p/z$ expresses how much higher the depth resolution becomes when compared to the spatial resolution of the system (e.g., $r = 2$ would mean 2x higher depth resolution than the spatial resolution, resulting in $z = 0.05$ mm/unitdisparity given $p = 0.1$ mm/pixel). All the parameters valid for our system are summarized in Table 3.1

In order to demonstrate how well our system performs for a high-quality printed matter (although this is not a metallic surface), images of a previously folded EUR-100 banknote where the absolute difference between the lowest and highest point was about 3 mm were acquired, see Fig. 3.19. One can see that the level of
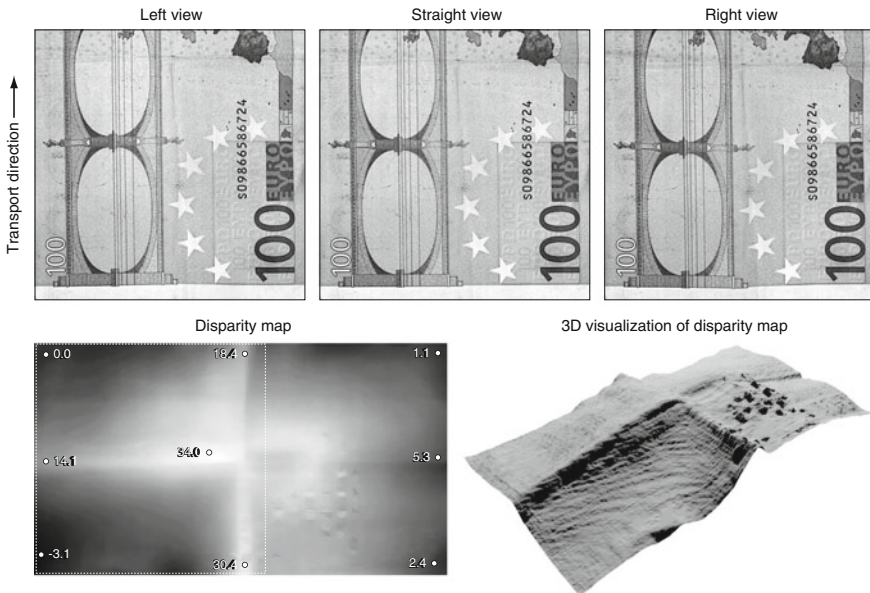
Fig. 3.19 Surface reconstruction of EUR-100 banknote: The upper three images show the three partial views. The *bottom left* image visualizes the disparity map estimated by the OpenCV SGBM approach. The overlaid numbers provide disparity estimates at the points marked. The 3D rendering of the disparity map is shown in the *bottom right* image

details stays about the same in all three views, so that even very fine details can be recognized. The disparity map estimated by the OpenCV SGBM approach shows folding artifacts of the banknote quite markedly. This effect is, however, even more clear in the 3D-rendered disparity map. In the bottom right segment of the disparity map, one may notice a strange regular structure of higher and lower disparity regions. This structure does not reflect any real relief details of the acquired banknote, but rather results from color moire artifacts occurred in that area in some of the views.

3.5.2 Line-Scan Light-Field Analysis

The multi-view system suggested in the previous section implements a specific light-field image acquisition system. We describe a more general approach to light fields in this section. Light-field capture and processing are methods of computational photography, an emerging field of research in computer vision. In general, computational photography is the combination of digital sensors, modern optics, actuators, and smart lights to escape the limitations of traditional film cameras, enabling for novel imaging applications [41]. We restrict ourselves to acquisition and processing of the light-field data in an industrial environment.

A light field is defined as the 4D radiance function of 2D position and 2D direction in regions of space free from occluders [29]. The two-plane parametrization, also called *light-slab* representation, of a light ray at position (u, v) and direction (s, t) is shown in Fig. 3.20. Formally, a light field $L(u, v, s, t)$ is a function which describes the amount of light passing through each point of space (u, v) (typically reflected from an object surface) in each direction (s, t). In practice, a light field captures each point of an object using a number of different light paths which allows for refocusing as well as changing parallax views.

The principle of light-field acquisition is compared to a standard setup as in Fig. 3.21, where Fig. 3.21a shows an in-focus portion of an object projected onto the image sensor plane. An out-of-focus image portion is projected to an extended
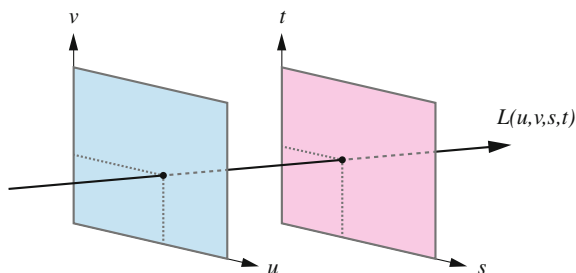
Fig. 3.20 Two-plane parameter representation of the light field: Each light ray is characterized by two spatial (u, v) and two directional coordinates (s, t)
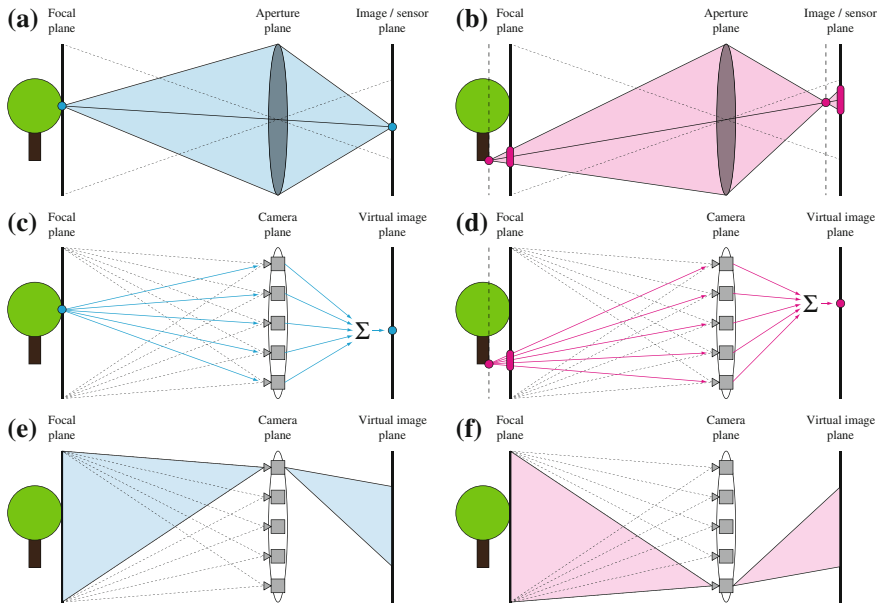
Fig. 3.21 Standard camera versus light-field camera principle: **a** Focused image in standard camera, **b** defocused image in standard camera, **c** partitioning of ray space and summing of light rays in light-field camera, **d** computational focusing in light-field camera, (**e**, **f**) viewpoint change in light-field camera

area, the so-called *circle of confusion*, using a standard camera, which is shown in Fig. 3.21b. Figure 3.21c depicts the light-field principle where virtual cameras are placed at positions (s, t) of the camera plane and an image is formed on a virtual image plane by summation of rays traveling from the object point (u, v) through all the virtual cameras. Figure 3.21d shows the summation of the light rays for an object portion lying not on the focal plane. Figure 3.21e, f shows so-called *sub-aperture images* which depict the object seen from two different viewpoints, thus defining a parallax view enabling for 3D interpretation.

Practical light-field acquisition can be realized in various ways. The simplest and bulkiest approach is a multi-camera array using a number of cameras capturing the same scene at the same time [62]. A gantry system uses a single camera which is mechanically displaced and a preferably still scene is acquired over time [29]. PiCam realizes a monolithic camera array using a 4 × 4 lens array placed on an equally organized low-resolution image sensor array [55]. An unstructured light-field acquisition using a moving hand-held device was also described recently [10]. Splitting the optical path using filters, masks, code patterns, etc. is termed *coded aperture imaging* [31]. Recently, the use of microlens arrays placed in front of the senor plane was realized in plenoptic cameras such as those manufactured by the companies Lytro and Raytrix, as decribed by Ng et al. [35], and Perwaß and Wietzke [38], respectively.

Our approach to light-field acquisition is tailored to typical requirements in industrial inspection, i.e., we consider a line-scan camera acquisition setup. In particular, an area-scan chip is used to capture object positions (u, v) under varying angle s over time. In this specific setup, there is no variation of the angle t measured across the transport direction. Therefore, we are dealing with 3D light fields (instead of 4D). Figure 3.22 shows the equivalence of an object line acquired by a number of line-scan cameras at once and a multiple line-scan camera observing a number of lines at one time and collecting multiple observations of an object line over time due to relative motion between camera and object. Note that this acquisition setup is also related to the TDI approach [63], with the difference that in TDI the object is acquired repeatedly under virtually the same angular direction and thus the computation reduces to a simple summation.

The data corresponding to varying observation angles at each line-scan sensor row are shown in Fig. 3.23, e.g., the object region A in the focal plane is seen under different angles indexed by s at different time indexes i. Summation along the diagonal line provides the irradiance estimate of the in-focus object region A. The data structure shown in Fig. 3.23 is called an epipolar plane image (EPI) when represented as a standard 2D image. EPIs were originally introduced for a structure from motion estimation [8], but they are still a popular tool in light-field data processing (e.g., [16, 26]).

It can be seen that parts of the object that do not reside in the focal plane of the camera result in a mapping to the sensor plane, which is different from the one described above. For such parts, corresponding radiance values map to EPI lines different from the diagonal, where the slope of the line depends on a distance between camera and that particular point of the object. Consequently, the summation in EPIs along non-diagonal slopes (usually requiring interpolation) allows for refocusing to virtual focal planes different from the main focal plane of the camera. Furthermore, this effect can be used the other way round, by detecting predominant orientations of EPI structures one can estimate the distance of the plane in which the particular object point is in focus. A combination of both, the EPI-based local estimation of predominant orientations, with the summation along
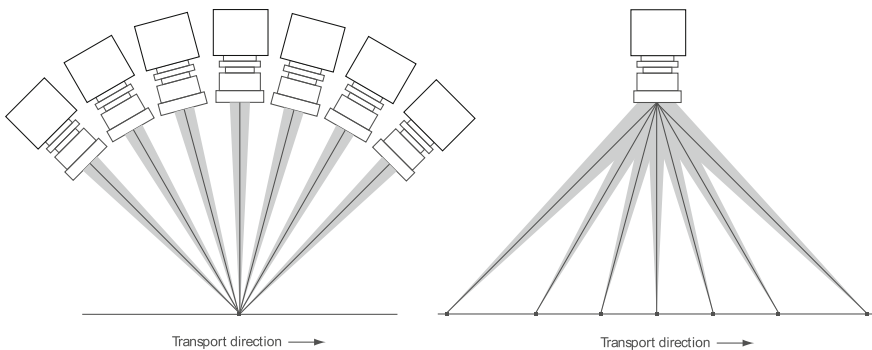
Fig. 3.22 Equivalence of many line-scan cameras and a light-field multiple line-scan camera

Fig. 3.23 Data structure acquired by a light-field multiple line-scan camera for object lines *A* through *G* over time index *i* and under different observation angles indexed by *s*
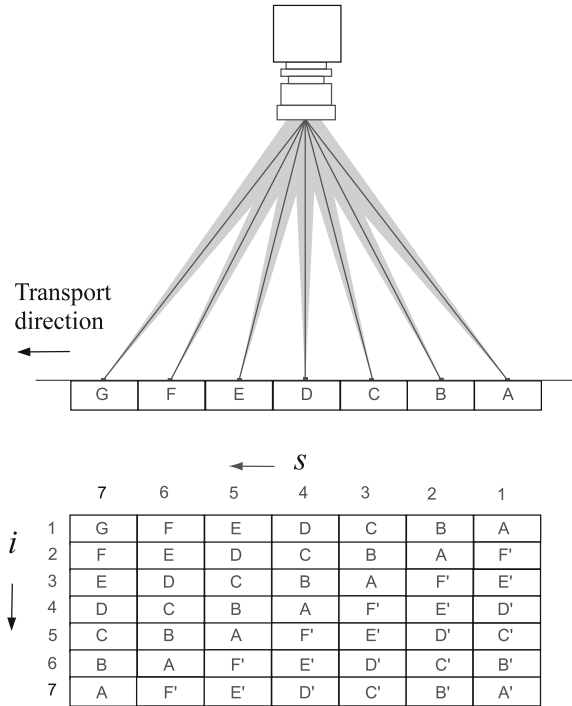
these estimated slopes, delivers the so-called *all-in-focus* image along with the depth map estimate.

Various strategies for the determination of predominant orientations of EPI structures exist. One frequently used approach is a so-called slope hypothesis testing. This method verifies a number of slope hypotheses in each location of the reference view (e.g., the central view). The winning hypothesis is then taken according to some criteria reflecting how well the slope matches with the pattern in the EPI domain. A common assumption is to assume Lambertian behavior of the scene, which presumes that all object points preserve their radiance values regardless of the viewing angle. Furthermore, the illumination is supposed to remain constant in different views of the light field. These rather strong assumptions are usually well preserved in natural scenes; however, they can be strongly violated in industrial applications, e.g., when dealing with highly specular metal surfaces. Let us now make a brief excursion into the state-of-the-art methods for depth estimation from light fields.

In [26], authors make the above two assumptions and, in return, they formulate an easy criterion for a good hypothesis, namely the best hypothesis is the one, where as many radiance values as possible along the hypothesized slope in EPI are similar enough to the radiance in the reference view. In other words, the winning hypothesis minimizes the overall deviation of radiance values along its slope with respect to the reference. Furthermore, they apply a so-called *fine-to-coarse*

refinement in order to complete unreliable depth estimates in homogeneous image regions by taking into account the coarser image scales. This approach has been shown to work very well under quite general natural scene conditions and can cope well with occlusions too. However, it is not so well suited for applications where one may expect non-Lambertian objects, non-constant illumination between the views, or higher camera noise.

Another approach, described in Venkataraman et al. [55], makes use of a pattern matching between different views for a discrete number of hypothesized depths based on the sum of absolute differences (SAD) of radiances. Also in this case, authors base their method to a certain extent on the mentioned two assumptions as SAD is a measure for the absolute difference between two patterns, although, providing greater robustness against outliers than the standard sum of squared differences (SSD). It should be noted that PiCam is a rigid 4×4 camera array, which captures 4D light field in 16 discrete points of the (s, t) space. One of the points is taken for the reference view, while all the others act on several predefined baselines with respect to the reference view. Similar to the previous approach, also PiCam performs well in natural scenes; however, its performance is disputable under more general conditions.

Last but not least, in Wanner and Goldlücke [58] authors take a statistical approach to the estimation of the principal orientation of EPI structures via analysis of the *structure tensor* constructed in each local neighborhood of EPI. This method effectively eliminates necessity of testing multiple hypotheses, which makes it well suited for high-performance applications; however, it again suffers from rather high sensitivity to noise, presence of non-Lambertian surfaces, non-constant illumination between the views, etc. In order to cover for this problem, the authors apply a global optimization (regularization) strategy, which significantly increase accuracy of resulting depth maps, however, at the expense of very high computational demands. Therefore, neither this method provides an ultimate solution to challenges posed by industrial applications.

Our approach to the determination of predominant slopes in the EPI domain uses the principle of slope hypothesis testing. It combines a pattern of cross-correlation measure with a measure of absolute deviation of radiances in a computationally efficient way, which makes it robust against disturbances such as noise, non-Lambertian surfaces, non-constant illumination. Furthermore, the method is capable of performing well with even very narrow-baseline systems, where precise subpixel disparity estimates are inevitable. In order to stabilize the results and provide better robustness against occlusions, we employ certain robust statistics such as SAD or MAD (the median of absolute differences) at several points of our algorithm. For more details on our approach, see [52].

Finally, we provide an example of a metal object acquired by our line-scan light-field approach—a copper coin. In Fig. 3.24, one can find the central view of the acquired 9-view 3D light field with a total angular difference between furthest views of only 1.2°. The camera operates at the spatial resolution of about 0.12 mm/pixel which gives the system the depth resolution of 5.73 mm/unit disparity if no subpixel disparity refinement is considered. Given the thickness relief

Fig. 3.24 Example of a metal
object acquired by our
multiple line-scan light-field
camera—a copper coin. The
figure shows the central view
of the acquired 9-view 3D
light field

details we are interested in, we are entirely dependent on an accurate estimation of
subpixel disparities making use of the light field. It should be noted that since the
coin exhibits a mixture of specular and diffuse reflection and furthermore the illu-
mination essentially travels with respect to the object in different views of our light
field, the appearance of some surface details may change significantly between the
views. This makes the subpixel disparity estimation an even more challenging task.

The resulting depth map generated by our hybrid pattern matching algorithm
from the coin light field is shown in Fig. 3.25. One can see that even very fine
surface details have been clearly revealed in the obtained depth map. Note that these
details are approximately one order of magnitude smaller than the estimated
physical depth resolution of our camera system.

As mentioned above, once the estimation of EPI slopes is finished, one can
generate the all-in-focus view by summation of radiance values along estimated
slopes in each object point. The resulting all-in-focus view obtained from the coin
light field is shown in Fig. 3.26a. When compared to the central sub-aperture image
and the synthetic full-aperture image (see Fig. 3.26b, c, respectively), the all-
in-focus image preserves focusing and details of the central sub-aperture image,
however, at the signal-to-noise ratio of the synthetic full-aperture image, which
already suffers from quite significant smearing artifacts due to a different speed of
motion at different distances. It should be noted that it is precisely this effect which
makes the standard TDI technology applicable to planar objects only. Our approach
enables the concept of TDI for non-planar objects by means of a more intensive
digital post-processing of data.

To conclude, in this subsection, we have discussed several light-field capture and
processing methods, which nowadays become a viable alternative to a traditional
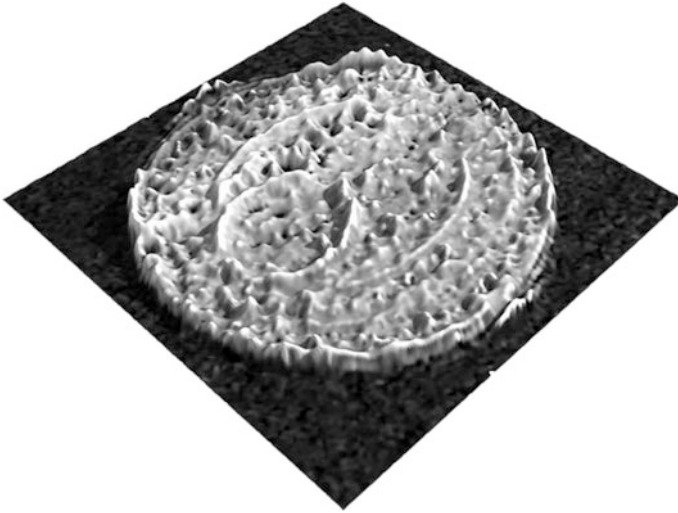
Fig. 3.25 3D depth model generated by our hybrid pattern matching algorithm from the coin light field. Note that even very fine surface details have been clearly revealed in the depth map

Fig. 3.26 Close-up views of the all-in-focus image (**a**), the central sub-aperture image (**b**), and the synthetic full-aperture image (**c**)

stereo-vision approach. We have also provided a real-world example of how these modern tools may help overcoming the limitations of traditional optics, cameras, illumination, and algorithms in challenging tasks of an industrial machine vision.

3.6 Conclusion

We have presented basic principles of reflection from metal surfaces and various acquisition setups operating in bright and dark field. These basic considerations led to the design of acquisition systems and analysis methods. The specific use and arrangement of light sources as well as redundancy is a key issue in the design of an

adequate acquisition system for industrial machine vision. Operating in dark field makes explicit use of the reflection properties of metallic surfaces, as these heavily change with illumination and surface orientation w.r.t. the image sensor as well as surface roughness. Detection of deviations from assumed reflectance behavior was exploited in rail surface analysis for specific disruption and textures. Stroboscopic lighting was used into overrule influences from the welding process itself, e.g., arc-glare, fume, and welding splatter. Finally, redundant observations from multi-view can, at least, reduce the influence of specular reflection on structured metal surfaces when operating in bright field.

In machine vision, algorithms for tracking, detection, classification, inspection, and reconstruction have to be selected based on a trade-off between accuracy and, most often, real-time capability. Furthermore, as one usually can control the environment in machine vision, the acquisition setup is usually designed in order to assist algorithms and reduce computational needs. To obtain real-time capabilities, the algorithm for rail surface disruptions bases its decision on efficiently computed local detection results. Similarly, quality assessment of weld beads is reduced to an efficient matching operation. Both tasks, rail and weld inspection, could have been approached also by more demanding classification approaches solved by machine learning, which was avoided.

Examples for more complicated tasks in machine vision of metal surfaces include optically variable metallic foil elements such as holograms and fine structured surfaces such as coins. We have presented an outlook how to analyze such data using multi-view and light-field imaging approaches. The suggested setups also address one peculiarity of machine vision, e.g., the use of line-scan acquisition.

References

1. Aoki K, Suga Y (1997) Intelligent image processing for abstraction and discrimination of defect image in radiographic film. In: Seventh international offshore and polar engineering conference, pp. 527–531
2. Aubreton O, Bajard A, Verney B, Truchetet F (2013) Infrared system for 3D scanning of metallic surfaces. Mach Vis Appl 24(7):1513–1524
3. Bae KY, Lee TH, Ahn KC (2002) An optical sensing system for seam tracking and weld pool control in gas metal arc welding of steel pipe. J Mater Process Technol 120(1–3):458–465

4. Basri R, Jacobs D, Kemelmacher I (2007) Photometric stereo with general, unknown lighting. Int J Comput Vision 72(3):239–257

5. Bennett H, Porteus J (1961) Relation between surface roughness and specular reflection at normal incidence. J Opt Soc Am (JOSA) 51(2):123–129

6. Biomedical Imaging Group: Local normalization (2002). URL http://bigwww.epfl.ch/demo/jlocalnormalization/ . Accessed 20 Dec 2013

7. Bishop CM (1995) Neural Networks for pattern recognition, chap. 2.6: Mixture models. Oxford Univerity Press, Oxford, pp. 59–73

8. Bolles RC, Baker HH, Marimont DH (1987) Epipolar–plane image analysis: an approach to determining structure from motion. Int J Comput Vis 1:7–55

9. Chen Z, Shi Y, Zhao H (2008) Evaluating technology of spot weld quality for coated high strength steel sheet based on ultrasonic guide wave. In: Ultrasonics symposium, IEEE, pp. 406–409

10. Davis A, Levoy M, Durand F (2012) Unstructured light fields. Comp Graph Forum 31 (2pt1):305–314

11. Dempster A, Laird N, Rubin D (1977) Maximum-likelihood from incomplete data via the EM algorithm. J Roy Stat Soc Series B (Methodological) 39:1–38

12. de Boor C (1978) A practical guide to splines. Springer, Berlin

13. Doherty A, Clark S, Care R, Dembowsky M (2005) Why rails crack. Ingenia 23:23–28

14. Fronthaler H, Croonen G, Biber J, Heber M, Ruether M (2013) An online quality assessment framework for automated welding processes. Int J Adv Manuf Technol 68(5–8):1655–1664

15. Gao X, Na SJ (2005) Detection of weld position and seam tracking based on Kalman filtering of weld pool images. J Manuf Syst 24(1):1–12

16. Goldluecke B, Wanner S (2013) The variational structure of disparity and regularization of 4D light fields. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 1003–1010

17. Grigorescu SE, Petkov N, Kruizinga P (2002) Comparison of texture features based on Gabor filters. IEEE Trans Image Process 11(10):1160–1167

18. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans Syst Man Cybern 3(6):610–621

19. Heber M, Lenz M, Rüther M, Bischof H, Fronthaler H, Croonen G (2013) Weld seam tracking and panorama image generation for on-line quality assurance. The Int J Adv Manuf Technol 65(9–12):1371–1382

20. Herwig J, Leßmann S, Bürger F, Pauli J (2013) Adaptive anomaly detection within near-regular milling textures. In: Ramponi G, Loncari c' S, Carini A, Egiazarian K (eds.) Proceedings of international symposium on image and signal processing and analysis, Trieste, Italy, pp. 106–111

21. Holländer B, Štolc S, Huber-Mörk R (2013) Multi-view line-scan inspection system using planar mirrors. In: Proceedings of SPIE optical metrology: videometrics, range imaging, and applications XII; and automated visual inspection, Vol. 8791. (Munich, D)

22. Hoßfeld M, Chu W, Adameck M, Eich M (2006) Fast 3D-vision system to classify metallic coins by their embossed topography. Electron Lett Comput Vis Image Anal 5(4):47–63

23. Hough PVC (1962) Method and means for recognizing complex patterns. In: U.S. patent 3,069,654

24. Huber-Mörk R, Nölle M, Oberhauser A, Fischmeister E (2010) Statistical rail surface classification based on 2D and 2.5D image analysis. In: Proceedings of international conference on advanced concepts for intelligent vision systems (ACIVS), Vol. 6474, Springer LNCS, Sydney, AU, pp 50–61

25. Jähne B (2002) Digitale Bildverarbeitung (5th edn), chap. 13.4.5: Quadraturfilter, Springer Berling, pp 383–384

26. Kim C, Zimmer H, Pritch Y, Sorkine-Hornung A, Gross M (2013) Scene reconstruction from high spatio-angular resolution light fields. ACM Trans Graph 32(4):73:1–73:12

27. Kim J, Son Y, Cho H, Koh K (1996) A robust visual seam tracking system for robotic arc welding. Mechatronics 6(2):141–163

28. Kuhl M, Neugebauer R, Mickel P (2007) Methods for a multisensorsystem for in-line process- and quality monitoring of welding seams using fuzzy pattern recognition. In: Emerging technologies and factory automation, pp 908–911
29. Levoy M, Hanrahan P (1996) Light field rendering. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, SIGGRAPH '96, ACM, New York, NY, USA, pp 31–42
30. Li G, Wang G, Zhong J (2004) A genetic algorithm on welding seam image segmentation. Intell Control Auto (WCICA) 3:2176–2178
31. Liang CK, Lin TH, Wong BY, Liu, C, Chen, H (2008) Programmable aperture photography: multiplexed light field acquisition. ACM Trans Graph 27(3):55:1–55:10
32. Myers CS, Rabiner LR (1981) A comparative study of several dynamic time-warping algorithms for connected word recognition. Bell Syst Tech J 60(7):1389–1409
33. Neuenschwander J, Blau B, Christin R, Luthi T, Rossler G (2004) Quality monitoring of the electron beam welding of the CMS conductor using ultrasonics. Appl Supercond 14 (2):526–529
34. Newman T, Jain A (1995) A survey of automated visual inspection. Comput Vis Image Underst 61(2):231–262
35. Ng R, Levoy M, Brédif M, Duval G, Horowitz M, Hanrahan P (2005) Light field photography with a hand-held plenoptic camera. Technical report, Stanford University
36. O'Gorman F, Clowes M (1976) Finding picture edges through collinearity of feature points. IEEE Trans. Computer C-25(4):449–456
37. Pang GK, Chu MH (2009) Automated optical inspection of solder paste based on 2.5D visual images. In: Proceedings of international conference on mechatronics and automation, pp 982–987
38. Perwaß C, Wietzke L (2012) Single lens 3D-camera with extended depth-of-field. In: Rogowitz BE, Pappas TN, De Ridder H (eds.) Proceedings of SPIE: human vision and electronic imaging XVII, vol 8291
39. Quanying D, Shanben C, Tao L (2005) Inspection of weld shape based on shape from shading. Int J Adv Manuf Technol 27(7–8):667–671
40. Rabiner LR, Juang B (1993) Fundamentals of speech recognition. Prentice-Hall
41. Raskar R, Tumblin J, Mpohan A, Agrawal A, Li Y (2006) State of the art report (STAR): computational photography. In: Proceedings of ACM/Eurographics. Vienna, AT
42. Rolfe J (1990) Optically variable devices for use on bank notes. In: Fagan WF (ed.) Proceedings of SPIE: Optical security and anticounterfeiting systems, 1210 (14):14–19
43. Sakoe H (1978) Chiba: Dynamic programing algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Sig Process 26(1):43–49
44. Samson V, Champagnat F, Giovannelli JF (2004) Detection of point objects with random subpixel location and unknown amplitude. Appl Opt 43(2):257–263
45. Sarosi Z, Knapp W, Kunz A, Wegener, K (2010) Detection of surface defects on sheet metal parts by using one-shot deflectometry in the infrared range. In: Proceedings of infrared camera applications conference 2010, Las Vegas, USA, pp 243–254
46. Schreiber D, Cambrini L, Biber J, Sardy B (2009) Online visual quality inspection for weld seams. Int J Adv Manuf Technol 42(5–6):497–504
47. Silva RD, Mery D (2007) State-of-the-art of weld seam inspection by radiographic testing: Part I: Image processing. Mater Eval 65(6):643–647
48. Silva RD, Mery D (2007) State-of-the-art of weld seam radiographic testing: Part II pattern recogtion. Mater Eval 65(8):833–838
49. Sonka M, Hlavac V, Boyle R (2008) Image processing, analysis, and machine vision (3rd edn), chap. 15.1.2: Co-occurrence matrices, Thomson, pp 723–725
50. Soukup D, Huber-Mörk R (2012) Cross-channel co-occurrence matrices for robust characterization of surface disruptions in 2.5D rail image analysis. In: Proceedings of international conference on advanced concepts for intelligent vision systems (ACIVS), vol 7517, Springer LNCS, pp 167–177
51. Stares I, Duffila C, Ogilvya J, Scrubya C (1990) On-line weld pool monitoring and defect detection using ultrasonic. NDT International 23(4):195–200

52. Štolc S, Huber-Mörk R, Holländer B, Soukup D (2014) Depth and all-in-focus images obtained by multi-line-scan light-field approach. To appear in: Proceedings of IS&T/SPIE Electronic Imaging—Image processing: machine vision applications VII. San Francisco, California, USA
53. Sun Y, Bai P, Sun H, Zhou P (2005) Real-time automatic detection of weld defects in steel pipe. NDT&E Int 38(7):522–528
54. Varma M, Zisserman A (2005) A statistical approach to texture classification from single images. Int J Comput Vis 62(1–2):61–81
55. Venkataraman K, Lelescu D, Duparré J, McMahon A, Molina G, Chatterjee P, Mullis R, Nayar S (2013) PiCam: An ultra-thin high performance monolithic camera array. ACM Trans Graph 32(5)
56. Wang G, Liao T (2002) Automatic identification of different types of welding defects in radiographic images. NDT&E Int 35(8):519–528
57. Wang Y, Sun Y, Lv P, Wang H (2008) Detection of line weld defects based on multiple thresholds and support vector machine. NDT&E Int 41(7):517–524
58. Wanner S, Goldlücke B (2012) Globally consistent depth labeling of 4D light fields. In: Proceedings of CVPR, pp. 41–48
59. Weman K (2003) Welding processes handbook. CRC Press, doi: 978-0849317736
60. Wentai Q, Yaoyao Z, Gangfeng Y (2005) Method of detecting welding seam based on the hough transform. Chinese J Mech Eng 41(9):212–216
61. Westin SH, Li H, Torrance KE (2004) A comparison of four BRDF models. In: Jensen H, Keller A (eds.) Proceedings of Eurographics symposium on rendering, pp 1–10
62. Wilburn B, Joshi N, Vaish V, Talvala EV, Antunez E, Barth A, Adams A, Horowitz M, Levoy M (2005) High performance imaging using large camera arrays. ACM Trans Graph 24 (3):765–776
63. Wong HS, Yao Y, Schlig ES (1992) TDI charge-coupled devices: design and applications. IBM J Res Dev Arch 36(1):83–105
64. Woodham RJ (1980) Photometric method for determining surface orientation from multiple images. Opt Eng 19(1):139–144
65. Xu P, Xu G, Tang X, Yao S (2008) A visual seam tracking system for robotic arc welding. Int J Adv Manuf Technol 37(1–2):70–75
66. Zacher M, Baumann G, Le R (2009) Modelle zur Prognose von Rollkontaktermüdungsschäden an Schienen. EI—Der Eisenbahningenieur 60(7):44–52
67. Zhang X, Xu J, Ge G (2004) Defects recognition on X-Ray images for weld inspection using SVM. In: Machine learning and cybernetics, 6:3721–3725
68. Zhang Z (2012) Microsoft Kinect sensor and its effect. IEEE MultiMedia 19(2):4–12

Chapter 4
FlexWarp, a Fast and Flexible Method for High-Precision Image Registration: A Nonparametric Approach for Precise Image Registration Applied to Industrial Print Inspection

Harald Penz, Franz Daubner, Ernst Bodenstorfer and Reinhold Huber-Mörk

Abstract We describe a fast nonparametric 2D image registration approach called FlexWarp. FlexWarp is designed for automated visual inspection. FlexWarp is discussed in detail, especially the algorithms for assessment and automatic selection of control points. A nonparametric image warping approach based on grids and image pyramids is explained. The capabilities of FlexWarp with respect to parallel execution are discussed. Implementation details on a GPU architecture and results are provided. The embedding into an industrial print inspection environment is discussed. Detailed results of FlexWarp behavior for print inspection and comparison to a state-of-the-art automatic registration method are presented.

Contents

H. Penz (✉) · F. Daubner · E. Bodenstorfer · R. Huber-Mörk
Intelligent Vision Systems, AIT Austrian Institute of Technology GmbH,
Seibersdorf, Austria
e-mail: harald.penz@ait.ac.at

4.1   Introduction

Image registration is an important operation in image processing. Automatic image registration is a prerequisite for detailed image comparison and inspection as well as a number of computer vision tasks such as optical flow computation, stereo matching, and visual tracking [54]. Technically speaking, image registration is often based on small image regions which could be identified well in corresponding images and later are used to establish a geometrical correspondence between images. Typically, each image region is associated with a point coordinate which refers to a position called a control point (CP). In the field of remote sensing, CPs are also sometimes called tie points and CPs were recently also referred to as interest points as the fields of applications became more widespread. Wide baseline stereo vision [37, 48], object recognition [36], and video indexing and retrieval [42] became new application areas for CPs.

We restrict ourselves to 2D image registration in industrial applications, where 2D tasks dominate 3D tasks. Registration of point clouds [16] and multi-view techniques [22] are beyond the topics of this chapter.

A typical registration procedure involving two images consists of the following basic steps

1. Detection of CPs in both images.
2. Matching of CPs between images, possibly involving consistency checks.
3. Mapping of one image to the other, sometimes also called warping.

Whereas in remote sensing, the manual or semiautomatic selection of CPs was dominating over decades [26], more automatic methods for tie point extraction have been reported recently also in remote sensing [34, 53] as well as in computer vision [46, 52]. Papers dealing with the problem of an automatic detection of CPs mostly concentrate on characteristic properties for such points, but some important details of the intended application are neglected. One such detail is related to the expected distortions. In the majority of papers, some parametric form of distortion is assumed, e.g., affine, perspective or simple translation, rotation, and scaling. Another important aspect in automatic CP extraction is the distribution, also called dispersion, of CP locations over the image, where often simple grid-based arrangements are used.

Most image registration procedures operate on pairs or small sets of images. In industrial inspection, a typical task is to register a large number of similar images, e.g., showing individual entities of the same product, at high speed. The second step

in registration procedures, i.e., matching and consistency checking, therefore becomes time-critical. Cross-validation of CPs between images and the wide-spread random sample consensus (RANSAC) [15] procedure for robust matching is characterized by high computational costs and variable run time and, therefore, becomes unfeasible in practice. In the suggested approach, CPs are extracted from the reference image only. In an off-line procedure, the CPs are assessed with respect to the measure of similarity and dispersed evenly over the image. In online execution, CPs are searched in comparison images. Instead of online consistency checking, we propose off-line assessment of CPs.

The assessment of CPs enables their automatic selection. Our assessment functions are adapted to the used measure of similarity in matching. Therefore, we introduce one assessment function for the normalized cross-correlation and another for the zero mean normalized cross-correlation. The main idea in the assessment functions is to explore the correlations of each CP with its neighborhood and the impact of distortions on the correlation values. Obtaining the assessment functions is computationally demanding and depends primarily on intended correlation window size and search range. The suggested method provides an estimation which correlation values could be expected. When matching CPs between images, this can be used to set a lower limit on the correlation value for each individual CP. CPs below this limit are regarded unreliable and should not be used for registration. Finally, selection of CPs is also guided by a procedure to improve dispersion of CPs by selection of evenly distributed CPs with high assessment values.

This chapter is organized as follows. Section 4.2 reviews image registration algorithms and published work. The FlexWarp approach is explained in detail in Sect. 4.3. Parallel implementation is described in Sects. 4.4 and 4.5 which shows integration in an industrial inspection environment. Experimental results exploring the algorithm itself and comparing it qualitatively with a state-of-the-art approach are presented in Sect. 4.6. Section 4.7 concludes this chapter.

4.2   Image Registration Techniques

CPs are usually defined as local image features where the image content varies significantly in more than one direction, e.g., at region corner locations. The requirements for interest points are high robustness with respect to partial visibility and high informational content. Therefore, it is argued that in this respect, interest points are superior to interest regions or edges [43]. Nevertheless, a number of region-based approaches to interest point detection appeared since then. For wide baseline stereo vision, the maximally stable extremal regions (MSERs) [37] and intensity extrema-based regions (IBR) as well as edge-based regions (EBR) [48] have been suggested. The most prominent interest point detectors are the one suggested by Harris and Stephens [21], the Hessian [28], the Moravec detector [39], and the one suggested by Förstner [18]. A method for detecting a set of interest points was suggested by Krattenthaler et al. [29], where only a subset of points

inside a region is used for matching based on correlation. Combinations of point and region detectors also appeared recently [51]. In order to detect interest points at different resolutions, image pyramid and scale-space approaches have been suggested. Among others, one of the most prominent approaches into this direction is the scale-invariant feature transform (SIFT) [36] and the speeded up robust features (SURFs) detector [4]. A scale-space detector of real-time performance is the center surround extrema (CenSurE) approach [1].

The most established detector for CPs is the second moment matrix. The second moment matrix A of averaged squared gradients, also known as the autocorrelation matrix, is widely used to describe the presence of local image structure such as corners. The matrix $A(x, y)$ for pixel position (x, y) is

$$A(x,y) = \begin{bmatrix} g(I_x^2(x,y)) & g(I_x(x,y)I_y(x,y)) \\ g(I_x(x,y)I_y(x,y)) & g(I_y^2(x,y)) \end{bmatrix}, \tag{4.1}$$

where g is local blurring function, e.g., a Gaussian smoothing operation. The eigenvalues of A provide information about the local image structure. Corner points are regarded as stable points if both eigenvalues are large. In order to avoid eigenvalue decomposition, the following measure of corner strength was suggested by Harris and Stephens [21]

$$R(x,y) = \det A(x,y) - k \operatorname{trace}^2 A(x,y), \tag{4.2}$$

where k is a constant, commonly chosen as $k = 0.04$.

Another well-known CP detector is based on the Hessian. The Hessian matrix H characterizing the neighborhood of an image pixel at position (x, y) is obtained by

$$H(x,y) = \begin{bmatrix} I_{xx}(x,y) & I_{xy}(x,y) \\ I_{xy}(x,y) & I_{yy}(x,y) \end{bmatrix}. \tag{4.3}$$

The corner strength can be derived from the matrix H by calculating its determinant [6].

The SIFT descriptor is based on gradient distribution in salient regions [36]. At each feature location, an orientation is selected by determining the peak of the histogram of local image gradient orientations. Subpixel image location, scale, and orientation are associated with each SIFT feature vector. Interest points are identified at peaks (local maxima and minima) of Gaussian functions applied in scale space, and all key points with low contrast or key points that are localized at edges are eliminated using a Laplacian function. SURFs are fast scale- and rotation-invariant features [4]. The SURF descriptor captures distributions of Haar wavelet responses within the neighborhood of an interest point. Multi-scale oriented patches (MOPS) are based on Harris corners and include a descriptor derived from scale space and adaptive non-maximal suppression [10]. A different approach also based on scale-space features is CenSurE. In CenSurE, center surround functions are approximated by bi-level filter masks efficiently computed by integral images

[12]. Scale invariance is accomplished by scaling the filter masks, non-maximal suppression is performed over the scale space, and line suppression is performed using properties of the second-order matrix, see Eq. (4.1).

Recently, for images containing homogeneous regions with well distinctive boundaries, MSERs were found to perform superior when compared to a number of other interest point detectors [38]. The method of MSERs uses a simple concept to delineate regions within an image [37]. An MSER is characterized by the property that its boundary is relatively stable over a wide range of pixel intensities; i.e., when performing operations with increased or decreased thresholds, there are only minor variations in the shape of the MSER.

The automatic selection of CPs for image registration is a topic of interest in remote sensing applications. A semiautomatic method started from a fixed grid of CPs is described in [27]. A method based on consistent labeling was described in [55], where so-called relational-based feature matches between multiple image pairs are extracted and refined by area-based matching. Assessment of optical flow estimation is related to CP assessment, and probability distributions of optical flow were studied on [45]. Selection of CPs based on affine image deformations and assessment of CPs for tracking purposes was discussed in [44]. The adaptive non-maximal suppression approach in MOPS aims at better spatial distribution of CPs, i.e., CPs are more uniform distributed over the image [10]. Automatic CP extraction based on SIFT and MOPS through consistent CP correspondence with respect to rigid 2D transformations is described in [41]. The work most similar to ours is dedicated to CP assessment for registration of remote sensing data [17], where CPs are assessed with respect to the quality of an estimated optical flow field between images. In their paper, a few methods to obtain an even dispersion of the CPs, i.e., a uniform spatial distribution of CPs, are compared, e.g., spatial subdivision and selection, spatial merging and the application of clustering based on k-means [35], and the LBG algorithm [33].

Once CPs are extracted, a measure of similarity is evaluated for pairs of potentially corresponding CPs. Point matching is typically performed by comparison of small subimages centered at the pixels in question in the considered images. Region and line matching approaches compare geometric properties such as shape, orientation, and dimensions. We consider point matching using image windows for comparison, also called template matching. Measures of match between template and target images range from mean absolute distance, mean squared distance, cross-correlation to normalized versions of correlation and information theoretical measures, such as mutual information. An overview of measures of match is given in [20].

The final step in registration is mapping one image to the other. Parametric mappings include 2D planar transformations such as translation, affine, and projective transformations. Parameters of these transformations are estimated from corresponding CPs. In the presence of outliers, robust estimation of parameters often uses RANSAC [15] or variants thereof. In case of non-rigid surfaces, one approach to map images is to use thin-plate splines (TPSs) [9]. TPSs are inspired by an analogy to the bending of a steel plate where the deviation from a flat surface is

used as analogy to planar displacement. Other nonparametric mapping methods use a deformable model for elastic deformation [40]. Such image deformations are commonly called rubber sheeting in the field of geographical information systems (GISs). Some rubber-sheeting approaches make use of triangulations, e.g. [19], while others use weighted interpolation to obtain displacement vectors in rubber sheeting [14, 31]. Our approach to rubber sheeting is similar to the latter ones.

4.3 FlexWarp Approach

FlexWarp uses a large number of control points in order to achieve high precision with respect to local registration. In order to achieve fast operation, the images are subdivided into a reasonable number of tiles. Most reliable control points are automatically selected in a preparation phase, which is a computationally demanding operation. Typically, the preparation phase is a less frequent operation performed off-line, i.e., not during online inspection where a reference print or master design image is compared to thousands of online acquired images. FlexWarp provides an assessment function for the selection of control points. The assessment considers not only the local neighborhood of the control point but also the size of the search range and even the correlation method. We show that slightly different control points should be used depending on the similarity measure, e.g., normalized or zero mean normalized cross-correlation. A flexible algorithm selects control points with high assessment values on the one hand and uniform spatial coverage on the other hand.

FlexWarp uses image pyramids to speed up control point search. At each pyramid layer, a preselected set of control points is searched. The sets of control points are independent between adjacent layers. Higher pyramid levels reduce the control point search in lower levels to small search ranges. Corresponding control points in the base pyramid level deliver displacement vectors distributed sparsely and non-uniformly over the image plane. Pixel-wise nonparametric displacement vector fields are derived from computed displacement vectors for a relative small number of points arranged on a regular grid. The grid partitions the images into rectangular tiles which are registered separately using the displacement vectors in the tile corners.

In image registration, typically either one image is chosen as reference image and CPs are selected in series of other images, e.g., in visual inspection or remote sensing applications, or series of images are pairwise registered in subsequent order, e.g., for tracking and optical flow applications. After proper CPs are selected in one image, they are searched in another image. This is usually done by some measure of similarity, with correlation being the most prominent one. Other measures of similarity include mutual information, distance functions, and divergence. Of course, the term CP is somehow misleading, as single points cannot be reasonable correlated, except for applications involving multispectral data. It is therefore common to use a surrounding area of the CP. From our point of view, a CP is not

just a point, but rather a small image section with a reference point. The reference point does not have an important meaning for us. It is needed because the result of a CP position search should be a point location for further computations. Commonly, the center of the image section is used as reference point. To decide whether an image section is suitable as a CP, following information should be taken into account:

1. The image section itself, i.e., its size and contents.
2. The intended search window.
3. The applied measure of similarity.

The first point is obvious and at least partially considered by most CP detection methods. A CP can be a bad choice if only a small neighborhood in the image is used and may become a good choice if a larger area is taken. In general, one can say that larger image sections are more robust but might be less accurate. As an example, look at the word MISSISSIPPI, where M is the only letter which can be used as CP because it is the only unique letter. If the search range covers the whole word, the letters I, S, and P cannot be used because they appear more than once. But combinations like PP and SIS are unique and can be used.

The second point is often neglected. If a CP is used with a large search window, some carefulness is required. There is a larger danger to mix it up with a similar image section nearby when compared to smaller search windows. The proposed letter groups in the MISSISSIPPI example above might be unusable if they appear in nearby text and the search window is large enough.

The third point is rarely considered, but different measures of similarity have different properties. While we admit that this fact is less important in practical examples, it should not be ignored. In summary, we suggest a two-step procedure for CP selection:

1. Calculation of an assessment measure for potential CPs.
2. Selection of CPs by a search strategy.

4.3.1 Measures of Similarity

An important issue in CP selection is the used measure of similarity. This issue is often neglected, e.g., there is no suggestion which measure of similarity is best used with Harris corners. We will shortly review the most established measures of similarity in this section before we will discuss the relationship between CP selection and measure of similarity in more detail in Sect. 4.3.2.

There are several well-known similarity measures [2]. We restricted ourselves to the normalized cross-correlation (NCC) and the zero mean normalized

cross-correlation (ZNCC). Given a reference image section P and a corresponding image I, the squared normalized cross-correlation is defined by

$$\text{NCC}^2(P, I, u, v) = \frac{\left(\sum_x \sum_y P(x,y) I(u+x, v+y)\right)^2}{\sum_x \sum_y P^2(x,y) \sum_x \sum_y I^2(u+x, v+y)}, \qquad (4.4)$$

where the offsets u and v denote a search vector. Due to computational reasons and to simplify the notation, the squared correlation value NCC^2 is used instead of NCC. The position of a CP is searched within a search window Ω:

$$\Omega = \{(u, v) | |u| \leq u_{\max}, |v| \leq v_{\max}\}. \qquad (4.5)$$

The search window Ω is a symmetric rectangular box as shown in Fig. 4.2a, which is the most common shape in image processing.

The squared zero mean normalized cross-correlation is defined by

$$\text{ZNCC}^2(P, I, u, v) = \frac{\left(\sum_x \sum_y P(x,y) - \overline{P}\big(I(u+x, v+y) - \overline{I}(u,v)\big)\right)^2}{\sum_x \sum_y P(x,y) - \overline{P}^2 \sum_x \sum_y \big(I(u+x, v+y) - \overline{I}(u,v)\big)^2}, \qquad (4.6)$$

where $\overline{P}$ is the average of the image section P and $\overline{I}(u, v)$ is the average of the corresponding part in the image I at search position (u, v). In order to enable one-pass calculation, Eq. (4.6) can be transformed to

$$\text{ZNCC}^2(P, I, u, v) = \frac{V^2(P, I, u, v)}{V(P) V(I, u, v)}, \qquad (4.7)$$

with

$$V(P) = n \sum_x \sum_y P^2(x,y) - \left(\sum_x \sum_y P(x,y)\right)^2, \qquad (4.8)$$

$$V(I, u, v) = n \sum_x \sum_y I^2(u+x, v+y) - \left(\sum_x \sum_y I(u+x, v+y)\right)^2, \qquad (4.9)$$

$$V(P, I, u, v) = n \sum_x \sum_y P(x,y) I(u+x, v+y) - \sum_x \sum_y P(x,y) \sum_x \sum_y I(u+x, v+y), \qquad (4.10)$$

where n is the number of pixels in the image section P. Using this formulation, there is no need to precalculate the average values $\overline{P}$ and $\overline{I}(u, v)$.

We assume that all pixel values are nonnegative. All NCC values will therefore be nonnegative too. But the ZNCC values can become negative and squaring them can be a problem. To avoid good matching with inverted images, we use

$$\text{ZNCC}^*(P, I, u, v) = \frac{V(P, I, u, v)|V(P, I, u, v)|}{V(P)\,V(I, u, v)}.$$ (4.11)

4.3.2   Assessment of CPs

A basic prerequisite for a CP selection algorithm is a numerical assessment for CPs and CP candidates, respectively. In this section, we will discuss the motivation for our assessment functions. Figure 4.1a shows three candidates for CPs. To simplify
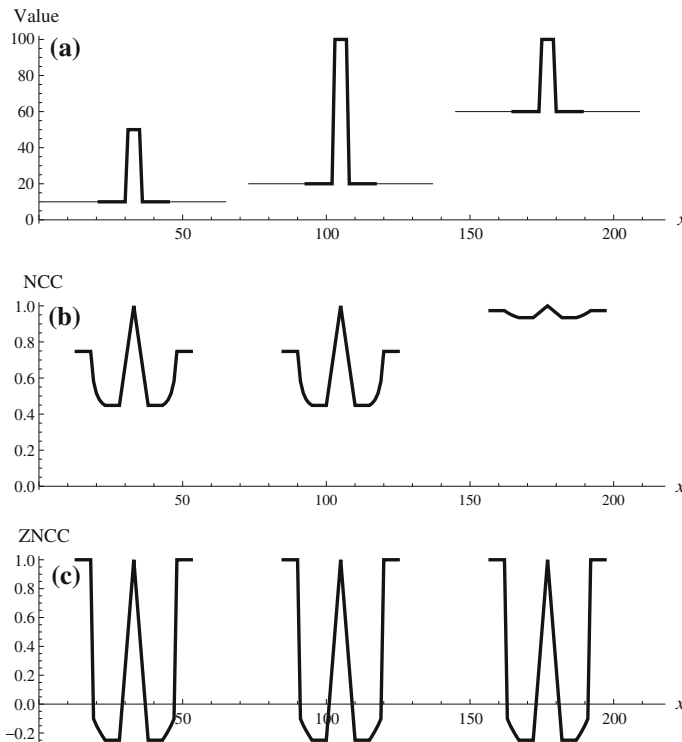
Fig. 4.1 Sample signals and their partial autocorrelation (PAC) graphs. **a** Three signals, only the thickened parts will be shifted and correlated. The *middle signal* is derived from the *left signal* with doubled amplitude. The *right signal* is derived from the *left signal* with a constant amplitude offset. **b** NCC functions for correlations between thickened parts and whole signals. **c** ZNCC functions for correlations between thickened parts and whole signals

matters, we have chosen one-dimensional signals instead of two-dimensional images for our example. Figure 4.1b, c shows the correlation functions for these three signals. Only the thickened parts of Fig. 4.1a are displaced and correlated. In Fig. 4.1a, the middle signal is derived from the left signal by doubling its values and the right signal is the left signal plus an offset in value. The corresponding NCC graphs are shown in Fig. 4.1b and ZNCC is used in Fig. 4.1c. Our goal is to find a formal expression for the assessment function for CP candidates. A human expert would quickly decide that the middle signal is best suited as CP, because it has the best contrast, while the shape is the same for all three signals. But it is not so easy to decide this in every case or give an assessment value for every sample.

We suggest that CPs should be assessed using the reference image only. The reason for this is that in several applications, e.g., typical inspection settings or any other online processing, there are no future observations available. We deduce our assessment functions from the correlation results of an image reference section. This partial autocorrelation (PAC) [PAC]PACPartial autocorrelation is similar to the well-known autocorrelation function. The autocorrelation function describes the cross-correlation of the full 1D signal (as in our example in Fig. 4.1) or 2D signal with itself for all possible or at least several relative shifts (often also called lags). It has to wrap around at the borders. PAC, however, correlates only a rectangular section of the image with the image. There is no wrap around at the borders. Correlation values are only computed for offsets where the shifted image section is completely overlapped by the image.

The NCC graph in Fig. 4.1b is the same for the left and middle signal. This is not what we would expect from a proper assessment function, as the middle signal should be preferred. The right signal shows a less distinctive PAC graph. This is fine to distinguish it from the middle signal, but it also distinguishes it from the left signal, although it is questionable that the left signal is a much better CP.

The PAC graphs for ZNCC in Fig. 4.1c are identical. This can be explained by the fact that ZNCC is designed to ignore additive and multiplicative variations. Figure 4.1c makes it clear that the PAC graphs alone are not sufficient to deduce a proper assessment function for CP candidates. We ignore this problem for the moment and discuss only the shapes of the PAC graphs. We always see a distinctive peak in the center. The top value of the peak is exactly 1 as NCC and ZNCC deliver this value if a signal is correlated with itself. That is true for all signals and does not help us to find good CPs. It is more interesting to look at the nonzero displacements of the PAC graphs. Our objective is to achieve considerably low correlation values in wrong CP positions. In other words, if there exists a significant displacement where the PAC delivers a high correlation value, the candidate signal is not suitable as CP. It could be mixed up with this similar section.

The peak in the middle of each PAC graph has significant decreasing slopes. But the graphs begin to increase for larger displacements. In the case of ZNCC, the PAC values reach the maximum value of 1 for larger shifts. We have to mention that this is not really a characteristic of ZNCC. The correlation value is actually not defined in these areas. It results in an expression of the form 0/0, and we decided to use the

substitute value 1. The problem comes from the fact that ZNCC cannot correlate a constant signal. The same problem arises with NCC if a signal is correlated with a zero signal. We used the correlation value 1 for such constellations to simulate a 'perfect' match. This marks the candidate as 'not unique' and prevents that it is chosen as CP. But that is only important if the search window is large enough to have the danger that the searched signal is correlated with the neighboring constant signal. The use of the substitute value 1 is probably a subjective decision because one can also catch these constellations during the position search of the CP and provide then a very low substitute value. This is what we actually do additionally. We therefore have two substitute values for constellations where the correlation fails: a high one (1 for NCC and ZNCC) in the CP selection phase and a low one (0 for NCC and −1 for ZNCC) during the CP position search. This is a double protection against extreme constellations.

4.3.3   Assessment Functions for CPs

We study the effects of small noise on the correlation coefficient in order to find a proper assessment function for CPs. Let P be an image section with n pixel. Since we do not care about coordinates in the moment, it is simpler to see P as a set of n arbitrary ordered pixel values p_i.

$$P = \{p_1, p_2, p_3, \ldots, p_n\}. \tag{4.12}$$

The pixels of P normally form a squared or at least rectangular image section. This makes an implementation easy. But our considerations are not limited to these shapes. Circles and even scattered pixels can also be used as CPs.

If we correlate P with itself, the correlation value is 1, which indicates a perfect match. Let us disturb one pixel of P by adding the value ε. Without loss of generality, we substitute p_1 by $p_1 + \varepsilon$ and call this modified image section P_ε.

$$P_\varepsilon = \{p_1 + \varepsilon, p_2, p_3, \ldots, p_n\}. \tag{4.13}$$

The squared NCC value of P and P_ε becomes

$$\mathrm{NCC}^2(P, P_\varepsilon) = \frac{(S_2 + \varepsilon p_1)^2}{S_2(S_2 + 2\varepsilon p_1 + \varepsilon^2)}, \tag{4.14}$$

where

$$S_2 = \sum_{i=1}^{n} p_i^2. \tag{4.15}$$

This is derived from Eq. (4.4), where i is substituted by P_ε and the offsets u and v are omitted. For fixed pixel values p_i, the function $\mathrm{NCC}^2(P, P_\varepsilon)$ depends only on ε and can be represented by the Taylor series

$$
\begin{aligned}
\mathrm{NCC}^2(P, P_\varepsilon) = 1 &- \frac{S_2 - p_1^2}{S_2^2}\varepsilon^2 + \frac{2p_1\left(S_2 - p_1^2\right)}{S_2^3}\varepsilon^3 \\
&+ \frac{\left(S_2 - 4p_1^2\right)\left(S_2 - p_1^2\right)}{S_2^4}\varepsilon^4 + O(\varepsilon^5)
\end{aligned}
\tag{4.16}
$$

Note that there is no linear term for ε. The exact impact of the disturbance ε depends on the pixel value p_1. The sum of squared pixel values S_2 is generally much larger than every squared single pixel value p_i^2. Alternatively, we can assume that $p_1 = 0$ and claim that the results for values $.p_1 \neq 0$ are not too different. We further assume that ε^2 is small compared to S_2 and neglect the higher order terms of ε. Equation (4.16) can then be simplified to the approximation

$$
\mathrm{NCC}^2(P, P_\varepsilon) \approx 1 - \frac{\varepsilon^2}{S_2}.
\tag{4.17}
$$

This expression shows that with larger S_2, higher values for ε are needed to produce significant smaller correlation values than 1. This seems to imply that CPs with more pixels and larger pixel values are preferable. But things are not that easy. We have to remember that we studied the impacts of disturbing just one pixel. It is obvious that the effects of changing only one pixel decrease with an increasing total number of pixels. It should not be concluded from this that more pixels are much better. Our model is not comprehensive enough for this conclusion. We are satisfied to find an assessment function which compares CP candidates of equal size.

Equation (4.17) shows that the impacts of deviations like noise on the correlation coefficient decrease with higher pixel values. But unfortunately, the effects on the correlation value produced by displacements may also decrease with higher pixel values. The rightmost signal in Fig. 4.1b shows this. High values for S_2 are therefore not enough.

We transform the approximation in Eq. (4.17) to

$$
\varepsilon^2 \approx \left(1 - \mathrm{NCC}^2(P, P_\varepsilon)\right)S_2.
\tag{4.18}
$$

A CP assessment function is a measure of the suitability of a CP candidate for matching purposes. CPs should be robust against perturbations. This is our main criterion for selecting CPs. Robustness means that large perturbations are necessary to influence the correlation coefficient stronger than a displacement would do. In a certain sense, we want to maximize Eq. (4.18). We want to have CPs where large perturbations are necessary to decrease the correlation value as much as a significant displacement would do. This leads us to the assessment function Q_{NCC} for NCC.

$$Q_{\text{NCC}}(P) = (1 - C_{\text{NCC}}(P, I_{\text{ref}}, \Omega_2))S_2, \tag{4.19}$$

with

$$C_{\text{NCC}}(P, I_{\text{ref}}, \Omega_2) = \max_{(u,v) \in \Omega_2} \text{NCC}^2(P, I_{\text{ref}}, u, v), \tag{4.20}$$

NCC^2 defined as in Eq. (4.4) and

$$\Omega_2 = \left\{(u, v)\,\big|\,|u| \le 2u_{\max}, |v| \le 2v_{\max}, u^2 + v^2 > r^2\right\}. \tag{4.21}$$

I_{ref} is the reference image from where the image section P is taken. To simplify matters, we assume that P and I_{ref} use the same coordinate origin. This means that $P(x, y) = I_{\text{ref}}(x, y)$ for all x and y. We use the doubled search window Ω_2 compared to the one used in Eq. (4.5) to make sure that a CP is not mixed up with a nearby similar image section. Figure 4.2 shows examples for the search window Ω and the larger window Ω_2. Search positions (u, v) are marked as dots. Search positions with a length smaller than or equal to r are removed in Ω_2. To understand why we use this larger window Ω_2, imagine an image with two very similar image sections and one of them is used as CP. If the position search is allowed to start in the middle of these two image sections, the search window must be large enough to reach the correct image section. But if the search window is symmetric, it would then be also large enough to reach the wrong image section. It is therefore advisable to use neither of these two image sections as CP. The doubled search window in Eq. (4.21) prevents this because Ω_2 includes then a vector to superimpose the two image sections. This results in a high value C_{NCC} and a low assessment function value Q_{NCC}.

The radius r regulates the minimal displacement vectors to be considered. We recommend rather small values like 2 or 3 pixels for r. We use $r^2 = 5$ for instance.

Analog considerations can be made for ZNCC. We obtain

$$Q_{\text{ZNCC}}(P) = (1 - C_{\text{ZNCC}}(P, I_{\text{ref}}, \Omega_2))\left(nS_2 - S_1^2\right), \tag{4.22}$$
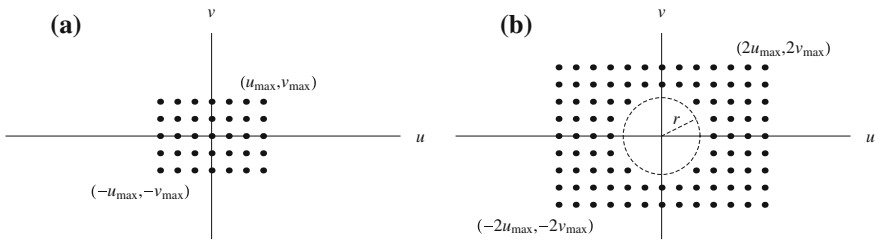
Fig. 4.2 Search windows. **a** Search window Ω. **b** Search window Ω_2, twice the size of Ω, but search positions smaller than or equal to r are omitted

where

$$C_{\mathrm{ZNCC}}(P, I_{\mathrm{ref}}, \Omega_2) = \max_{(u,v) \in \Omega_2} \mathrm{ZNCC}^*(P, I_{\mathrm{ref}}, u, v), \qquad (4.23)$$

ZNCC^* defined as in Eq. (4.11), S_2 defined as in Eq. (4.15), and

$$S_1 = \sum_{i=1}^{n} p_i. \qquad (4.24)$$

It can be seen that the assessment function for ZNCC in Eq. (4.22) is similar to the assessment function for NCC in Eq. (4.19). The sum of squared pixel values is replaced by a more complex term which is basically the pixel value variance, a measure for the image contrast.

Q_{NCC} and Q_{ZNCC} are our proposed assessment functions for the suitability of image sections as CPs, one for NCC and a different for ZNCC. The assessment function values are computed from a single image. This makes the method especially feasible in practical applications, because more images are sometimes not available or the effort to manage and use them is not achievable in practical applications. We strongly recommend using a fixed size for the image sections because the assessment result for different sizes is not comparable.

Table 4.1 lists the experimentally obtained values for the signals shown in Fig. 4.1. The radius r as shown in Fig. 4.2b has the value 2. The maximum search distance u_{max} has the values 4 and 8, respectively. $v_{\mathrm{max}} = 0$ because the signals are one-dimensional. As suspected, the assessment function values Q for the middle signal of Fig. 4.1a are much higher than for the others. The zero values for Q_{ZNCC} indicate that the search distance $u_{\mathrm{max}} = 8$ was wide enough that correlations with a constant signal appeared. ZNCC does not give a useful result in this case and the substitute correlation value 1 is used, which forces Q_{ZNCC} to be 0.

Table 4.1 Experimental results for the signals shown in Fig. 4.1 with $r = 2$ and $v_{\mathrm{max}} = 0$

	$u_{\mathrm{max}} = 4$			$u_{\mathrm{max}} = 8$			
	Left	Middle	Right	Left	Middle	Right	
S_1	450	900	1700	450	900	1700	Equation (4.24)
S_2	14,500	58,000	122,000	14,500	58,000	122,000	Equation (4.15)
C_{NCC}	0.669	0.669	0.961	0.747	0.747	0.973	Equation (4.20)
C_{ZNCC}	0.250	0.250	0.250	1.000	1.000	1.000	Equation (4.23)
Q_{NCC}	4800	19,200	4800	3663	14,650	3243	Equation (4.19)
Q_{ZNCC}	120,000	480,000	120,000	0	0	0	Equation (4.22)

4.3.4 Lower Limits for CP Correlation Coefficients

It is often hard to decide whether the position found by a CP should be used or not. The highest found correlation value denotes the final position and might be used as indicator whether an individual CP is suited for matching. Unfortunately, CPs can develop very different behavior, even within the same image. For example, a correlation value of 0.9 might be sufficient for one CP and useless for another CP. Therefore, we propose an individual lower limit with respect to each CP P $L_{\mathrm{NCC}}(P)$ based on its correlation value. We derive this individual limit L_{NCC} from the PAC.

$$L_{\mathrm{NCC}}(P, I_{\mathrm{ref}}, \Omega_2) = \kappa + (1 - \kappa) C_{\mathrm{NCC}}(P, I_{\mathrm{ref}}, \Omega_2). \tag{4.25}$$

C_{NCC} is defined as in Eq. (4.20) and the factor κ is a user-defined parameter. The parameter κ has to be smaller than 1. Although it is not completely out of question, κ is normally greater than 0. We recommend $\kappa = 0.5$ for NCC. For ZNCC, even lower values like $\kappa = 0.25$ seem to be useful. Again, this depends of course on the intended purpose.

It is possible to deduce a kind of probability or quality function for CP results. We suggest to use the following function

$$G_{\mathrm{NCC}}(P, I, I_{\mathrm{ref}}, \Omega, \Omega_2) = \frac{C_{\mathrm{NCC}}(P, I, \Omega) - C_{\mathrm{NCC}}(P, I_{\mathrm{ref}}, \Omega_2)}{1 - C_{\mathrm{NCC}}(P, I_{\mathrm{ref}}, \Omega_2)}. \tag{4.26}$$

This function provides some feedback about the 'quality' of a CP position result. The value of this function does not exceed 1. A value of 0 means that the correlation coefficient is as low as the largest possible correlation coefficient of the PAC when only displacements larger than the radius r of Eq. (4.21) are considered. The found CP position should be ignored if $G_{\mathrm{NCC}} < \kappa$, the same parameter κ as in Eq. (4.25).

It is often proposed to study the consistency of the found positions of CP sets and eliminate outliers. The definition for outliers can be refined with our quality function G_{NCC} and thus improved the image registration. For example, an outlier P is only discarded if G_{NCC} is small.

4.3.5 Complexity Estimations

The computation of the assessment of a CP requires the analysis of all possible displacements in Ω_2 of Eq. (4.21). The size of Ω_2 has the order

$$|\Omega_2| = O(u_{\max} v_{\max}). \tag{4.27}$$

The time complexity $T_{\text{assess}}(P, \Omega_2)$ of the assessment function for the image section P with n pixels and the displacements in Ω_2 has the order

$$T_{\text{assess}}(P, \Omega_2) = O(nu_{\max}v_{\max}). \tag{4.28}$$

We want to assess each possible image section P with a given shape and size. The assessment time complexity of the full reference image I_{ref} with width w_{ref} and height h_{ref} can be estimated in the order of

$$T_{\text{assess}}(I_{\text{ref}}, P, \Omega_2) = O(w_{\text{ref}}h_{\text{ref}}nu_{\max}v_{\max}). \tag{4.29}$$

An optimization is possible, provided the image section P is a rectangle, which is fortunately most usual. This assumption makes the assessment effort independent from the size n of the image sections and brings the time complexity $T_{\text{rect}}(I_{\text{ref}}, \Omega_2)$ for the reference images I_{ref} with rectangular CPs down to

$$T_{\text{rect}}(I_{\text{ref}}, \Omega_2) = O(w_{\text{ref}}h_{\text{ref}}u_{\max}v_{\max}). \tag{4.30}$$

The presented optimization is based on the fact that the assessed image sections are strongly overlapping. The computations for an image section and its neighboring image section, which is only one pixel displaced, are largely identical, which can be used in a beneficial way. The sums S_1 and S_2 of Eqs. (4.24) and (4.15) can be efficiently computed with summed area tables (also known as integral images), as introduced by Crow [12]. The memory requirements of these tables might be a problem if large images must be assessed. We prefer a recursion formula to compute the sums of rectangular image sections:

$$\sum_{x=a}^{b}\sum_{y=c}^{d}I(x,y) = \sum_{x=a-1}^{b-1}\sum_{y=c}^{d}I(x,y) - \sum_{y=c}^{d}I(a-1,y) + \sum_{y=c}^{d}I(b,y), \tag{4.31}$$

$$\sum_{y=c}^{d}I(x,y) = \sum_{y=c-1}^{d-1}I(x,y) - I(x,c-1) + I(x,d). \tag{4.32}$$

This method has the same time complexity as the summed area table method, but needs less temporal memory. Only the column sums $\sum_{y=c}^{d}I(x,y)$ must be temporary stored, which requires the storage of only one data line, as these data can be overwritten for each new line.

4.3.6   CP Selection Algorithm

The assessment functions Q_{NCC} and Q_{ZNCC} of Sect. 4.3.3 provide us with an overview, which areas are more suitable for CPs than others. Our current goal is an

algorithm capable of selecting spatially evenly distributed CPs. In a first step, we reduce the number of potential CPs by one or more of the following methods:

1. Discard all potential CPs P with an assessment value $Q(P)$ below a minimal value.
2. Discard all potential CPs P whose assessment value $Q(P)$ is not a local maximum.
3. Separate the reference image I_{ref} into disjoint parts (e.g., 10×10 pixel patches) and discard all but the best potential CP from each part.

We call the remaining list of potential CPs as candidate list, denoted by L_{Cand}. From this candidate list L_{Cand}, we create a CP list L_{CP}. Starting with an empty set, the CPs are iteratively added to L_{CP}. Each newly added CP should have a large distance to all other CPs. We define the squared distance $\delta^2(P, L_{\mathrm{CP}})$ between the CP P and the CP list L_{CP} as

$$\delta^2(P, L_{\mathrm{CP}}) = \begin{cases} w_{\mathrm{ref}}^2 + h_{\mathrm{ref}}^2 & \text{if } L_{\mathrm{CP}}=\emptyset, \\ \min_{P_{\mathrm{L}} \in L_{\mathrm{CP}}} \left((x(P) - x(P_{\mathrm{L}}))^2 + (y(P) - y(P_{\mathrm{L}}))^2 \right) & \text{otherwise,} \end{cases} \tag{4.33}$$

where $x(P)$ and $y(P)$ denote the coordinates of the reference point of P. w_{ref} and h_{ref} are the width and height of the image I_{ref}. Any constant value larger than 0 is regarded to be a suitable default value for the case that L_{CP} is empty. But the expression $w_{\mathrm{ref}}^2 + h_{\mathrm{ref}}^2$ ensures that $\delta^2(P, L_{\mathrm{CP}})$ is always monotonic decreasing when new CPs are added to L_{CP}, even for the first CP.

On the one hand, a CP should have large distances to all other CPs; on the other hand, a CP should have a high assessment value. Therefore, we define the following rating function $f(P, L_{\mathrm{CP}})$ in order to find a compromise between these two demands:

$$f(P, L_{\mathrm{CP}}) = \delta^2(P, L_{\mathrm{CP}}) Q^{\gamma}(P). \tag{4.34}$$

We use either Q_{NCC} from Eq. (4.19) or Q_{ZNCC} from Eq. (4.22) for the assessment function Q. The exponent γ enables us to adjust the relation between the two demands.

Since we normally use the first method of the above list to reduce the number of potential CPs, a lower limit for the assessment values $Q(P)$ is already installed. We use therefore low values like $\gamma = 0.3$ which emphasize more on the first demand of large distances between the CPs. If the assessment quality is more important, higher values for γ should be used.

Algorithm 1 shows the details. The CP list L_{CP} is empty at the beginning. $\delta^2(P, \emptyset)$ is equal for all P, and $f(P, \emptyset)$ only depends on the assessment value $Q(P)$. This means that the first selected CP is the one with the largest assessment value. The candidate list L_{Cand} remains unchanged over the selection procedure. The fact that $\delta^2(P, L_{\mathrm{CP}}) = 0$ and $f(P, L_{\mathrm{CP}}) = 0$ if $P \in L_{\mathrm{CP}}$ prevents a renewed selection of a

CP. The condition $\delta^2(P, L_{CP}) \geq \delta_{min}^2$ in line 5 is optional. This condition prevents that two CPs are selected which have a distance below the given limit δ_{min}. The algorithm stops when the desired number l_{max} of CPs was found. Alternatively, the algorithm may stop before this if no further suitable CPs are available. This early stopping might occur, when the candidate list L_{Cand} was too small or δ_{min} was chosen too large.

The time complexity $T_{select}(L_{CP}, L_{Cand})$ to compute the CP list L_{CP} from the candidate list L_{Cand} is determined by the two nested loops in Algorithm 1.

$$T_{select}(L_{CP}, L_{Cand}) = O(|L_{CP}||L_{Cand}|). \tag{4.35}$$

Algorithm 1 Select a list L_{CP} of up to l_{max} of evenly distributed CPs from the CP candidate list L_{Cand}. The distance between every two selected CPs must be greater than or equal to δ_{min}. The rating function $f(P, L_{CP})$ determines which CPs are selected.

```
1:  L_CP ← ∅
2:  repeat
3:      f_best ← 0
4:      for all P ∈ L_Cand do
5:          if δ²(P, L_CP) ≥ δ²_min and f(P, L_CP) > f_best then
6:              f_best ← f(P, L_CP)
7:              P_best ← P
8:          end if
9:      end for
10:     if f_best > 0 then
11:         L_CP ← L_CP ∪ {P_best}
12:     end if
13: until |L_CP| = l_max or f_best = 0
```

The computation of the rating value $f(P, L_{CP})$ in line 5 of Algorithm 1 seems to require an additional time factor $O(|L_{CP}|)$ as P must be related to all CPs in the list L_{CP}. But the list L_{CP} changes only by one additional element in each iteration of the repeat loop. Instead of a complete recalculation, it is more efficient to store the squared distance value $\delta^2(P, L_{CP})$ for each $P \in L_{Cand}$ and update these values with each iteration of the repeat loop by the recursion formula

$$\delta^2(P, \emptyset) = w_{ref}^2 + h_{ref}^2, \tag{4.36}$$

$$\delta^2(P, L_{CP} \cup \{P_{best}\}) = \min\left(\delta^2(P, L_{CP}), (x(P) - x(P_{best}))^2 \right.$$
$$\left. + (y(P) - y(P_{best}))^2\right) \tag{4.37}$$

which can be derived from Eq. (4.33).

4.3.7 Registration

We devised a method to select proper CPs from a reference image in the previous sections. This gives a set of shift vectors. If we use the lower limits of Sect. 4.3.4, we have also the information which CPs and corresponding deviations can be used. The final step is to make use of these deviations in the image registration.

One approach would be to compute an affine transformation, for example, with a least square fit approximation. A more flexible approach is to use the reference points of the CPs as input for a Delaunay triangulation [13]. Every triangle can be assigned to a unique affine transformation determined by the corner points.

Our approach is probably simpler to implement. After searching the CPs in the sample image and discarding CPs with poor results, we have a set of arbitrary arranged points. Every image point location obtains a tentative shift vector indicating the displacement between reference and sample image. In the next step, we transform the point coordinates into a completely new set of regular arranged grid points. The reference image is divided into rectangles of equal size, and the corners of these rectangles form a new regular arranged grid point set. We compute the shift vectors for these grid points. One method to do this is to use thin-plate splines (TPSs) where shifts are related to the physical analogy of bending a thin sheet of metal. TPS is a method with good quality but high computational effort, at least if the number of CPs is large. We decided to use a rubber-banding approach, which uses the much simpler inverse distance weighting (IDW) approach. In our implementation, for each grid point, a small set of nearest CPs is selected. Any missing shift vector is computed with a weighted sum from these neighbors. In IDW, the weights are reciprocal to the distance and the sum of the weights is 1. This could be applied to each pixel rather than only to the grid points. Unfortunately, this could become computationally costly as well.

It is more efficient to use the regular arranged grid points to compute the pixel shift vectors. In this approach, every pixel uses the 4 neighboring grid points which are the 4 corners of the surrounding rectangle. A linear interpolation is used to compute the x- and the y-component of the shift vectors as shown in Fig. 4.3 and Eq. (4.38).

$$v = \frac{(a - a_x)(b - b_y)v_{0,0} + a_x(b - b_y)v_{a,0} + (a - a_x)b_y v_{0,b} + a_x b_y v_{a,b}}{ab}. \quad (4.38)$$

These pixel shift vectors v for their part are used to interpolate the pixels.

4.3.8 Image Pyramids

We suggest to use image pyramids to speed up and improve the registration procedure [11]. Using this well-known approach, the reference image and a sample image are repeatedly smoothed and subsampled to build the pyramids. Every

Fig. 4.3 Vector interpolation: the vector v at position (a_x, b_y) is a linear combination of the vectors $v_{0,0}$, $v_{a,0}$, $v_{0,b}$, and $v_{a,b}$
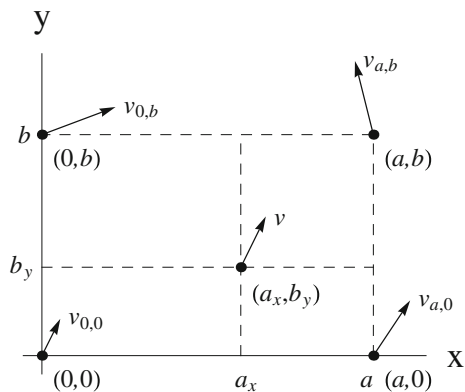
pyramid level has its own set of CPs (See Fig. 4.4). We normally use CPs of constant size (measured in pixel) for all pyramid levels. Since the number of pixel decreases exponential with increasing pyramid height, we use significantly less CPs for higher levels. Note that the CP selection process for each pyramid level is independent from the other pyramid levels.

Location of CPs in images drawn from an unseen sample becomes more difficult. First, the CPs of the top pyramid level are located in the sample image. This gives a rough estimation of the transformation to be used for the next lower pyramid level.
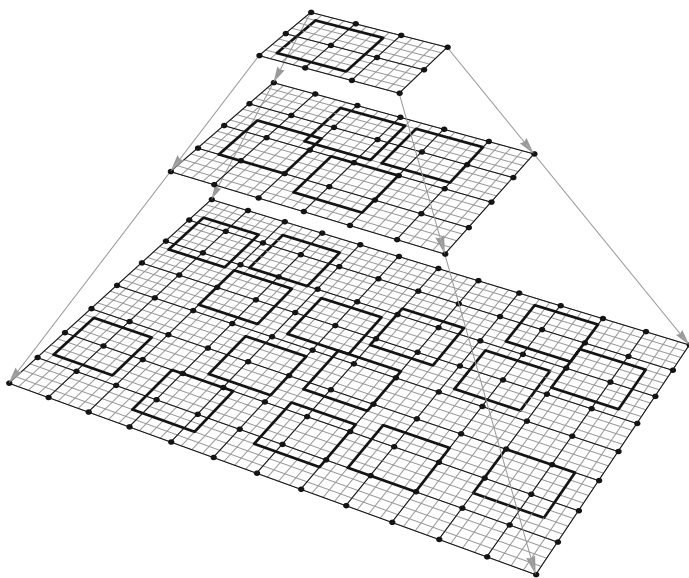
Fig. 4.4 Example for a reference image pyramid: *Bold squares* are CPs, and *dots* indicate grid positions

The image parts required for the CP search are registered with this initial transformation and propagated to the next level, and specific CPs are located again. This procedure is continued down to the bottom level of the image pyramid. This procedure has several advantages. The most important is that only small search windows are needed. The CP search is faster and the results are more reliable. As pointed out in Sect. 4.3.3, the number of CP candidates increases when the search window is smaller.

More CP candidates mean greater choice and a potential improvement in quality. The correlation methods NCC and ZNCC are sensible to distortions. We therefore use the information from the coarser pyramid level to register the area where a CP for the actual pyramid level is searched. The distortions are reduced, and the found CP positions are more accurate.

Each pyramid level exploits the information of the coarser level, except for the topmost pyramid level. FlexWarp is designed to be used in conjunction with other methods. The idea is to apply first a less accurate but more robust method—against large distortions—and feed the top pyramid level with this information. This can be seen as special level at the top of the pyramid.

A less obvious advantage of the image pyramid is that it provides us with acceptable replacement values if a CP fails. We have not mentioned in Sect. 4.3.7 if the neighboring CPs for each grid point are selected online or off-line. We do this actually off-line. Since CPs might fail, we need replacement values and vectors, respectively, because we want to avoid an online search for available CPs which can be used instead. These replacement shift vectors are taken from the overlying pyramid level. We simply omit the failing CPs and transfer their summation weights to the shift vector proposed by the overlying pyramid level. If all CPs for a grid point fail, this proposal is taken unchanged, i.e., with weight 1.

4.4 GPU Implementation

Since their introduction, there has been a remarkable increase in the performance and capabilities of GPUs. The modern GPU is not only a powerful graphics engine but also a highly parallel programmable processor featuring peak arithmetic and memory bandwidth that substantially outperforms its CPU counterpart. This usage is referred to as GPGPU (General-purpose computing on graphics processing units). The downside to this speed increase is that the algorithm needs to be suitable for massive parallelism, since GPUs are built around fundamentally different philosophies than traditional CPUs.

Our final goal would be a full FlexWarp, i.e., including off-line and online processing, implementation on a GPU. To get started, we decided to address the most time-critical part, which is the last part of the online processing. The actual rectification is the computationally most demanding part of the online processing, and it is very well suited for GPU implementation. Furthermore, the first part of

online processing can be tweaked for more speed by choosing less CPs, or using an appropriate combination of pyramid levels. Nevertheless, the cost of the actual rectification is dictated by the image size, which usually cannot be altered for a given application. The rectification part can be applied independently to each pixel, which comprises a lot of independent tasks. This is a promising prerequisite for good GPU performance. On the other hand, we have an irregular access pattern to the source image data, as the displacement of each pixel may be different, which might have a negative impact on memory throughput.

Our implementation platform consists of a quadcore CPU and a Nvidia Geforce GTX 285. Since the graphics card is rather dated by today's standards, we also tested the final implementation on a system with a Geforce GTX Titan. The CPU reference implementation is written in Standard C++, with the most time-critical parts parallelized to run on all four CPU cores. Apart from the parallelization, the CPU version is not further optimized. For most algorithms, taking advantage of the specialized vector processing units included in modern CPUs (SSE2, AVX, etc.) can yield a significant performance improvement. In this case, however, optimizations with vector units would be quite tricky because of the nonlinear access pattern to memory. We assume that careful optimization on the CPU yields a factor 3–4 speed improvement.

In order to obtain a measure of the maximum attainable memory bandwidth on the GPU, we performed a series of measurements. Due to published datasheets, Nvidia claims the maximum memory bandwidth of our test card (GTX 285) is 159 GB/s, where 1 GB = 10^9 bytes. This number does not include overheads, so we expected our measured results to be lower. As there is no direct equivalent of a CPU memcpy function within a GPU program, we wrote a short kernel which transfers the data with parallel read and write accesses. From the twofold data block size (read + write) and the kernel run time, we are able to estimate the data throughput. We measured the results with different access widths per thread and with different sizes of data blocks to be transferred. For the results of our measurements, refer to Table 4.2. In the optimal case, our test system achieved a throughput of 130 GB/s, which is about 20 % less than the nominal 159 GB/s. This performance is well within our expectations. The maximum throughput was achievable only by choosing the optimal access width per thread and with a sufficiently large data set. If the data set is not as large or an unfavorable number of bytes per thread is read, the performance drops significantly. Note that we always used the optimal memory

Table 4.2 Measured GPU memory throughput depending on access width and size of data set

Access width (bytes)	Performance (GB/s) for different data size				
	1 MB	4 MB	16 MB	64 MB	256 MB
1	11.8	19.3	21.5	22.4	22.8
2	17.2	31.2	42.9	47.4	48.5
4	23.1	58.6	84.1	109.4	115.2
8	23.7	64.2	100.4	120.1	130.0
16	20.4	53.6	88.7	104.2	109.9

access pattern, where every thread accesses the memory location directly after the previous thread. With a random access pattern, the performance was at least an order of magnitude lower.

When implementing the rectification kernel in the online part of FlexWarp, we decided to start with a simple approach. We reformulated the algorithm in a way that it was applicable to each pixel individually, but we did no further optimizations. This first GPU version was already a bit faster than the CPU version, but speedup was not yet convincing. We occupied the GPU with a sufficient number of threads, which is a good choice, but the memory access pattern was far from optimum performance.

We were able to achieve a little speed increase by preloading the displacement vectors in the so-called shared memory. With this modification, the vectors no longer were loaded separately for each pixel. Instead, the vectors were loaded only once for each pixel block. The main problem, however, which was the access to the actual image data, remained. The source image, i.e., the image before rectification, needed to be accessed four times in an irregular pattern for each rectified output pixel.

The breakthrough came when we started to leverage the texturing units. Since the standard task of a GPU is the rendering of complex 3D graphics, it includes a series of specialized hardware units for that purpose. Some of that functionality can be used with CUDA, for example, a small cache for the access to the GPU memory and bilinear interpolation. It turns out that these texture units are a perfect fit for the problem at hand. The resulting kernel is much shorter in code than previous versions, since most of the work is done in hardware. All read accesses (vector data and source image) are performed by the texture units, and the bilinear interpolation is done on the fly while reading from memory. This means that instead of the 4 memory accesses needed for bilinear interpolation, we only issue one read instruction to the texturing unit. The interpolation of the 4 neighboring pixels is done transparently in hardware, utilizing a cache to reduce bandwidth requirements. Due to cache coherence, the irregular access pattern is no longer impeding performance either.

For our results with different image sizes, see Table 4.3. We included measurements with the CPU reference implementation, with our test GPU (GTX 285), and with the more modern GTX Titan. For an image size of 4000 × 4000 with four color channels (64 MB), we measured a data throughput of around 23 GB/s on the GTX 285. Compared to the performance of the reference implementation, we achieved a speedup of 279×. If we account for the fact that it might be possible to make the CPU version faster by a factor of four, we would still get a speedup of 70×.

Table 4.3 Measured total throughput for rectification performed on CPU and GPU

Image size	1 MB	4 MB	16 MB	64 MB	256 MB
CPU	0.060	0.071	0.079	0.084	0.085
GTX 285	10.204	17.002	21.702	23.420	23.704
Titan	6.221	15.359	34.407	42.800	50.397

We also measured the performance of our implementation on a Nvidia GTX Titan, unfortunately in combination with a different (slower) CPU host system. The GTX Titan has a nominal memory bandwidth of 288 GB/s, which is about 80 % faster than the GTX 285; additionally, it has a lot more streaming processors and texturing units. With the largest data set, we see more than a 100 % speed increase on the GTX Titan, which suggests that our algorithm scales nicely with the newer hardware. With the smallest data set, we have almost a 40 % drop in performance. The actual difference in run time is only 0.065 ms. Since we have to measure the run time of the program with the CPU system, the performance drop might have been introduced by the slower host system of our Titan card.

We should mention that until now, we have assumed that the image data already reside in GPU memory when the rectification operation starts. Note that if we have to transfer it first from the CPU main memory, we have to add that time. If, however, several images should be rectified, the process can be pipelined, since transfer from the CPU to GPU memory and calculations on the GPU can be done in parallel. That way, we are limited only by the slower of both processes. In the considered case, the limiting factor would be the PCI-express bus transfer rate (of approximately 5.5 GB/s). This leaves room for additional calculations on the GPU, if necessary.

4.5 Integration into Industrial Inspection Systems

In this section, the practical application of the FlexWarp algorithm in print inspection is illustrated. The intention is to give a better understanding of various causes for image distortions, the compensation of which FlexWarp was originally designed for, and to address issues arising from high-speed requirements.

In Fig. 4.5, we show main elements of a typical print inspection system for sheets of paper, e.g., packages of consumer products, banknotes, or stamps. The printed matter is transported by means of a conveyor belt in front of one (or several) digital camera(s). Hence, line-scan cameras can be used for image acquisition. Examples for state-of-the-art line-scan cameras are [3, 47]. The acquired image data are fed into the image processing system (IPS), which computes the decision to

Fig. 4.5 Operational diagram of a typical print inspection system
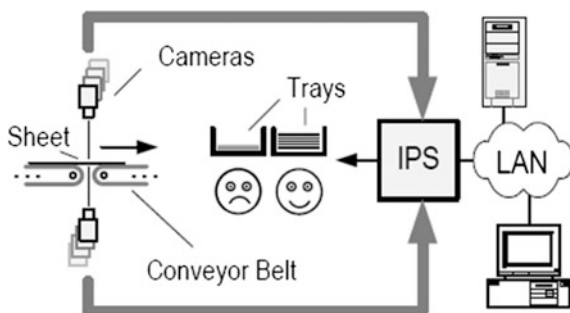
separate waste from high-quality products in real time. The main requirements for image acquisition in print inspection are nowadays

- acquisition of color images,
- high line rates for maintaining high throughput of the printing process, and
- fine resolution for appropriate sensitivity to detect faint flaws.

Today's print inspection systems are designed for machine speeds in the range of 10 m/s scanning of printed matter with a typical pixel resolution of 100 μm or less. To achieve this, all components of the inspection system, covering lighting, camera lenses, cameras, high-speed data links, and the image processing system (algorithms and hardware) have to be selected and designed carefully for proper collaboration at high speed. Assuming a 2k pixel color line-scan camera yields an image data rate of 100,000 lines per second × 2000 pixels × 3 color channels = approx. 600 MB/s. This means that the camera/data link/image processing system has to be fast enough to deliver/transmit/process this amount of data (per camera).

Trying to make print inspection systems faster and resolution finer, one faces a couple of problems that have to be solved.

1. Data rates explode,
2. When resolution becomes finer, modeling image distortions as a global affine transform is not accurate enough,
3. Exposure times decrease; lighting power cannot be increased any more; and therefore, camera noise becomes a limiting factor.

As for the problem of exploding data rates, in most cases, parallelization of high-speed interconnects and image processing hardware will be the answer. In Sect. 4.4, an example for parallelization using a GPU is given.

The second problem is due to the fact that coming down to resolutions of few micrometers induces that local distortions of the inspected sheet during transport become relevant. Examples for effects leading to local image distortions are flutter effects, slip, vibrations, elastic deformation, lens distortion, or imperfections of the line trigger signal generation (e.g., shaft encoder). FlexWarp is the solution we propose. FlexWarp was invented for compensation of those small local image deformations to be able to exactly register the acquired image with a master image.

The third problem is one of the main challenges for the design of high-speed image sensors and especially high-speed line-scan sensors. For example, a line rate of 100 kHz means that the exposure time has to be less than 10 μs. The exposure time defines the time interval, when the electronic shutter of the sensor is open and photogenerated charges are collected within the pixels. Therefore, high-speed line-scan applications require especially bright lighting modules and camera lenses with high aperture. Otherwise, video signal levels would be too low and image quality would suffer. So, to raise line rates beyond the state of the art, image sensors have to become both faster and more light-sensitive, which are in general conflicting requirements in electronic circuit design.

As described in the article by He and Nixon [23], a solution to this dilemma is time delay integration (TDI). A TDI sensor is an image sensor suited for line-scan

applications where light is the limiting factor. Instead of having only one line of pixels scanning the object, there are multiple lines of pixels arranged in a matrix, thus capturing the object several times. The number of lines is referred to as number of TDI stages. In general, the number of lines (i.e., number of rows of the pixel matrix) is much less than the number of pixels per line (i.e., number of columns of the pixel matrix). Summing up multiple (N) exposures of the same object has the advantage that the signal to noise ratio (SNR) is increased by $\sqrt{N}$, if N denotes the number of TDI stages. The $\sqrt{N}$ law is very generally valid. The only assumption is that the shot noise coming from the quantum nature of photogenerated charges is the dominant noise source in the signal processing chain. This is true in most cases. Exceptions are mainly the dark parts of an image, where the video signal is far below saturation. For the dark image parts, the increase in SNR depends on the exact implementation of the TDI process. A good overview over the different implementations of TDI can be found in [32]. Depending on where the accumulation process takes place, there are three types of TDI:

- Accumulation in the charge domain,
- Accumulation in the voltage or current domain, or
- Accumulation in the digital domain.

Traditionally, all TDI sensors have been implemented as charge-coupled devices (CCDs), belonging to the first type. CCD sensors allow to shift photogenerated charges across the sensor. Therefore, CCD sensors are predestined for implementing TDI. If charges are shifted synchronously with the object's movement, i.e., with same speed and in the same direction as the projection of the object on the sensor surface, charges are automatically accumulated while being shifted. The charge accumulation process in CCD is nearly perfect, because it does not introduce additional noise, which is essential for very low-light applications, e.g., acquiring fluorescent objects. A drawback of CCD TDI is that an electronic shutter is difficult to implement, meaning that the exposure time of a CCD TDI sensor depends on the object's speed, leading to brightness changes in the image when transport speed changes. As compared to image sensors based on complementary metal oxide semiconductor (CMOS) technology, in most cases, manufacturing of CCDs is more expensive, operation voltages and power consumption are higher, and line rates are lower. Therefore, in the last years, there were several attempts to exploit the TDI effect using CMOS image sensors. The advantage of the CMOS fabrication process is that additional electric circuits such as column-parallel analog to digital converters (ADCs) or low-voltage differential signaling (LVDS) outputs [30] can be easily implemented on chip, which is essential for reaching high line rates. Another advantage is that the rows of the pixel matrix can be accessed randomly. Therefore, the exposure time can be fixed although the transport speed changes, and pixel rows coated with different color filters are possible because accumulation is not limited to signals of directly neighboring rows. This enables multi-linear color filter layouts which are more robust against Moiré effects than approaches with different color filters on top of the pixels of different columns, e.g., due to a Bayer filter pattern [5].

Fig. 4.6 Enhancement of responsivity measured at CMOS TDI sensor [8] is proportional to number of TDI stages
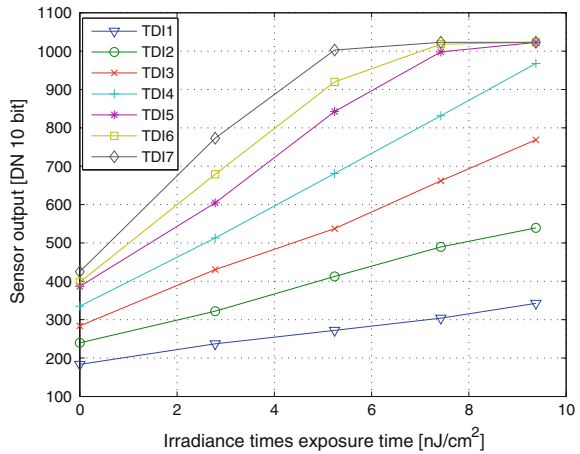
A comparison of a multi-line color CMOS TDI sensor, developed at Austrian Institute of Technology, with an image sensor with Bayer filter pattern is shown in [8], proving the better suppression of Moiré artifacts for the multi-line color pixel arrangement. This CMOS TDI sensor implements TDI as accumulation of voltages in analog circuit design. A maximum line rate of 125 kHz and a maximum number of 8 TDI stages were reached. Figure 4.6 shows the measured sensor transfer functions for different numbers of TDI stages (TDI1,…, TDI7), documenting the effect of increase in light sensitivity. Note that for the 5, 6, and 7 TDI stages, responsivity is so high that saturation (clipping) occurs within the range of measured irradiance values.

An example for TDI in digital domain is [7]. An existing CMOS active pixel area image sensor was used to acquire multi-line data. The TDI accumulation was done off-chip, in a field-programmable gate array (FPGA). Color TDI was possible when using a variant of the sensor with Bayer filter pattern. Based upon this principle, the first color line-scan camera known to the authors that was capable of delivering 100,000 color lines per second was designed. The number of TDI stages was two. This digital TDI technology has already been practically applied in a print inspection system.

The described advances in high-speed line-scan cameras, in highly accurate image registrations with the FlexWarp algorithm, and in parallelization of image processing algorithms form the foundation for future high-performance inspection.

4.6   Experimental Results

In the following, we present qualitative results for FlexWarp and a comparison to a state-of-the-art automatic method for image registration. The considered data are images from machine vision, i.e., the inspection of printed matter and from book scanning.

Fig. 4.7 Printed newspaper image used to assess the properties of FlexWarp

4.6.1 Properties of FlexWarp

Figure 4.7 shows an image scanned from an Austrian newspaper page. We use this as a reference image from which CPs should be generated. In the following discussion we will concentrate on the grayscale subimage of size 615×320 pixel which is shown in (Fig. 4.8).

Figure 4.9 shows the reference image with 100 CPs, marked by black squares. The CPs were found with our method. The size of the CPs is $n = 16 \times 16$ pixel. We used NCC as measure of similarity and $u_{max} = v_{max} = 2$ pixel as search range.

Fig. 4.8 Reference image

Fig. 4.9 Reference image with marked CPs. NCC, $n = 16 \times 16$, $u_{max} = v_{max} = 2$

Figure 4.10 shows the same section but from another exemplar of the newspaper. We use this as sample image, in which the CPs of the reference image should be found. The sample image is slightly larger (639×344 pixel) than the reference image to make sure that every CP is within the image.

Figure 4.11 shows an erroneous attempt to register the sample image in Fig. 4.10 to the reference image in Fig. 4.8 by using the CPs from Fig. 4.9. The errors occurred because the search range $u_{max} = v_{max} = 10$ was used, while for the CP generation, only the search range $u_{max} = v_{max} = 2$ was taken into account. This

Fig. 4.10 Sample image

Fig. 4.11 Erroneous sample image registration. CPs for a too small search area were used. NCC, $n = 16 \times 16$, $u_{max} = v_{max} = 10$ for image registration, but $u_{max} = v_{max} = 2$ was used in the CP generation process. The *red* marked CPs were found at wrong locations (color figure online)

difference has led to 3 CPs with wrong coordinates. Two of these CPs are on the left side: one covering the clock time 18 and the other covering the clock time 20. The third erroneous CP is in the last newspaper column and covers the clock time 19. Due to the similarity of printed text with respect to lines these 3 CPs where found shifted up by one text line, which resulted in local distortions. The image size of the registered image is identical to the image size of the reference image. The black squares marking the CPs are drawn at the same positions as in Fig. 4.9 as these are the target positions, where the CPs should end up.

Figure 4.12 shows a successful image registration. This time the required search range $u_{max} = v_{max} = 10$ was used for generating the 100 CPs. The CPs are now chosen in a way that they are unique within the search range thus avoiding problems as in Fig. 4.11.

Figure 4.13 shows the assessment function in the form of an image. This assessment function was used to produce the CPs of Figs. 4.9 and 4.11, respectively. Dark areas indicate good positions for CPs, whereas bright areas should be avoided. No assessment values are calculated for the border pixel because we need space for the CP and the vectors in Ω_2 of Eq. (4.21). The width of the white border is therefore half the size of a CP plus the doubled search range. These are $\sqrt{n}/2 + 2 * u_{max} = 16/2 + 2 * 2 = 12$ pixel. We applied the square root on the assessment values to improve the contrast. The values $\sqrt{Q_{NCC}}$ were scaled. The maximum value was mapped to 1 (white) and 0 was mapped to 0 (black). Finally, the image was inverted thus mapping the largest value of $\sqrt{Q_{NCC}}$ to 0.

Figure 4.14 shows the assessment function used to generate the CPs of Fig. 4.12. It looks similar to Fig. 4.13. One significant difference is the wider white border. This comes from the larger search range. Another significant difference is that some specific areas are much brighter than in Fig. 4.13. These are areas where the patterns seemed to be good CPs but are not unique within the full search range. Larger search ranges generally lead to smaller but never higher assessment values, and one could

Fig. 4.12 Sample image registration with marked CPs. NCC, $n = 16 \times 16$, $u_{max} = v_{max} = 10$
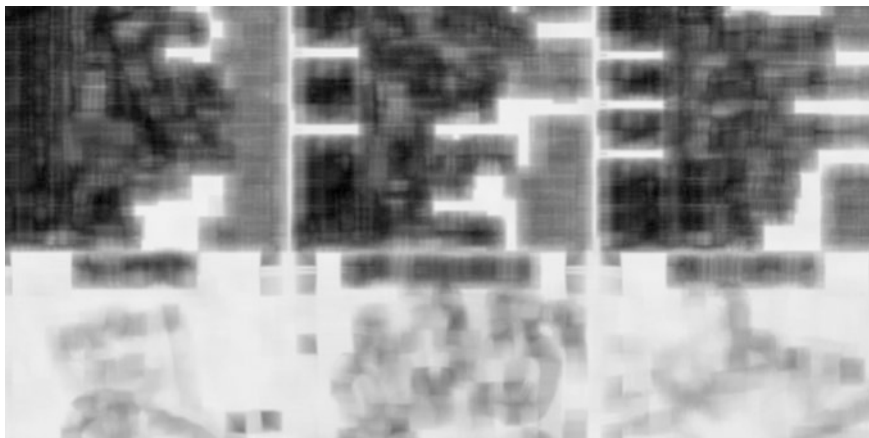
Fig. 4.13 CP assessment image. NCC, $n = 16 \times 16$, $u_{max} = v_{max} = 2$
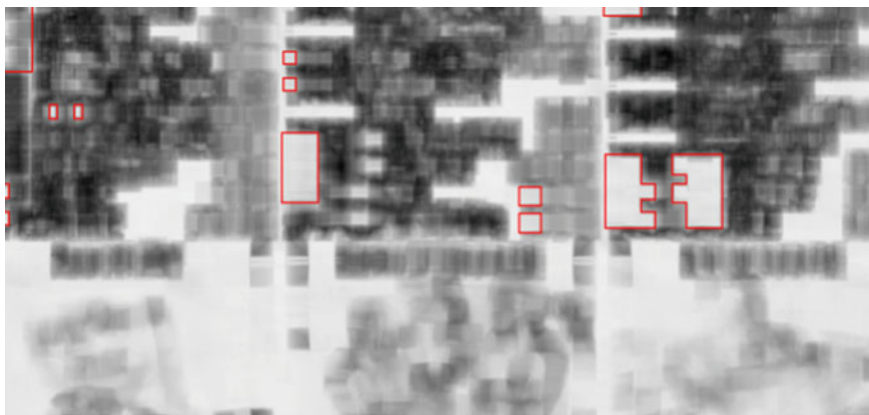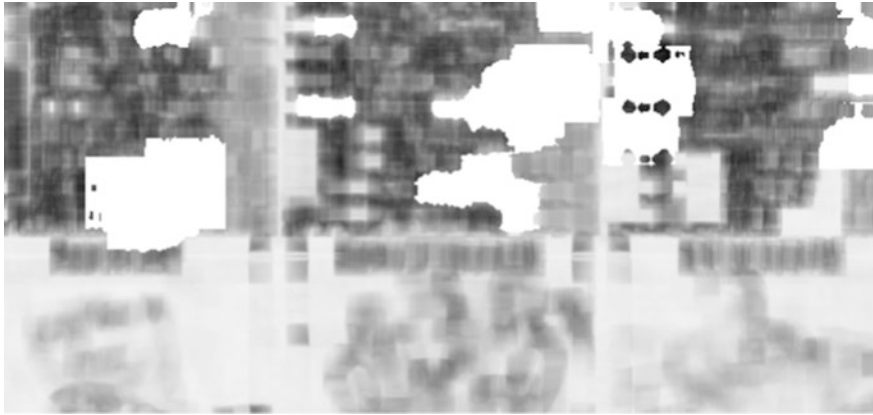

Fig. 4.14 CP assessment image. NCC, $n = 16 \times 16$, $u_{max} = v_{max} = 10$. Some bright areas which do not appear in Fig. 4.13 are marked with *red borders*. They are not suitable for CPs with the larger search area (color figure online)

therefore expect that each pixel in Fig. 4.14 should be at least as bright as in Fig. 4.13. But we used the same method for the brightness as for Fig. 4.13, and because the largest value of Q_{NCC} is different, we have a slightly different brightness scaling.

Figure 4.15 shows the assessment function if CPs for the measure of similarity ZNCC should be used. This image is similar to Fig. 4.14 because besides the different correlation method, the same parameters were used. But there are large white areas. These are caused by larger pieces of white paper. During the computation of the assessment function, a term close to 0/0 occurred and was captured. This happens if a CP is correlated with a homogeneous or almost homogeneous section.

Fig. 4.15 CP assessment image. ZNCC, $n = 16 \times 16$, $u_{max} = v_{max} = 10$

Fig. 4.16 CP assessment image. NCC, $n = 32 \times 32$, $u_{max} = v_{max} = 10$

Figure 4.16 shows the assessment function for CPs of the size $n = 32 \times 32$. The other parameters are the same as for Fig. 4.14. The larger CP size caused a reduction of the bright areas. But the white border where no assessment values could be computed increased due to the larger CPs.

4.6.2   Comparison to a State-of-the-Art Automatic Registration Approach

We compare the FlexWarp approach to the widely used SIFT approach for automatic CP selection applied to affine image registration [36]. In the latter approach, selection of CPs is based on robustly fitting an affine homography between master

Fig. 4.17 Images taken from scanned book data used to compare the FlexWarp to a SIFT/affine image registration method

Table 4.4 Comparison of FlexWarp and SIFT/affine image registration

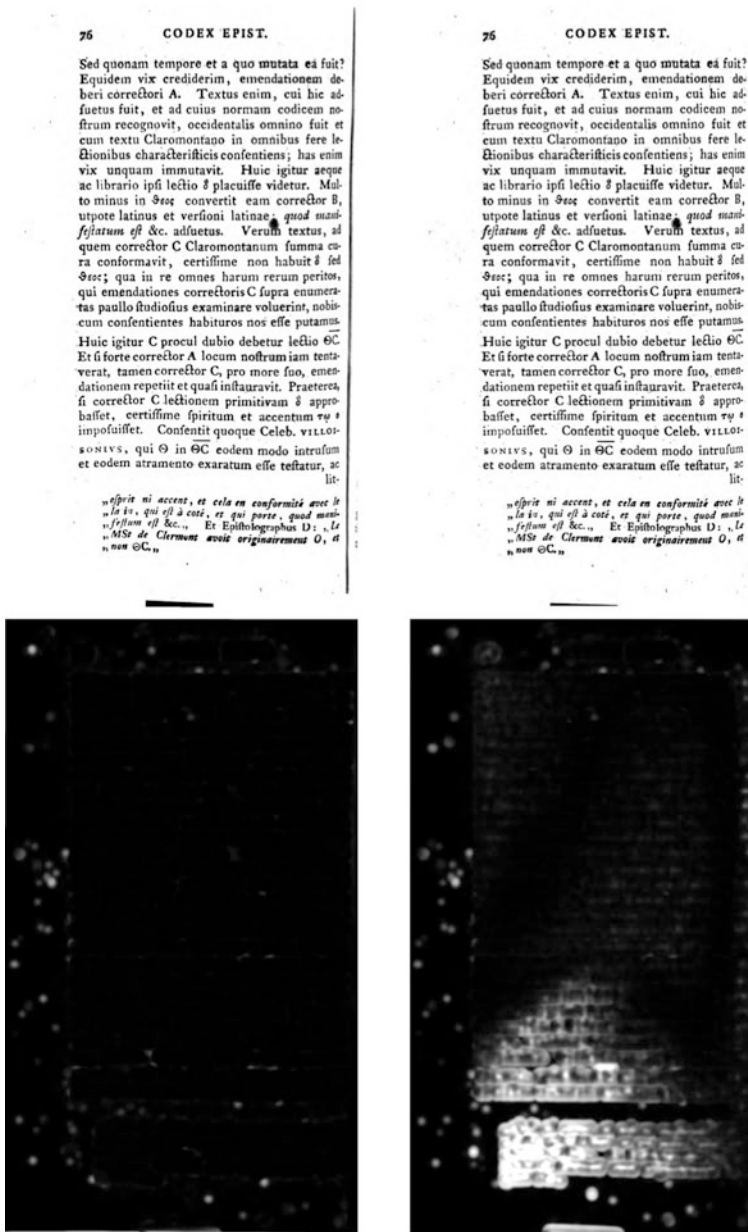|	FlexWarp			SIFT/affine		
Image name	MSSIM	MSE	MAD	MSSIM	MSE	MAD
Liber fecundus	0.98	73.09	2.98	0.98	141.95	3.50
Liber primus	0.98	117.29	4.78	0.98	134.30	4.82
Hieronymus	0.98	263.04	6.54	0.96	805.89	10.52
Tabak	0.99	170.72	4.51	0.99	216.94	4.42
Codex epist	0.99	192.91	5.25	0.90	569.58	13.19
Statistik	0.98	217.39	3.86	0.98	338.53	4.86

Fig. 4.18 Corresponding pairs of images (*upper row*) and SSIM images for registered images using FlexWarp (*lower left*) and for the SIFT/affine method (*lower right*): *Black color* indicates low SSIM and *white color* refers to high SSIM

and sample images using the RANSAC procedure [15]. We investigate data obtained by a large-scale scanning project of historical books, where automated visual inspection of duplicated pages and detection of page distortions is an issue [24, 25].

The comparison of methods is based on the measurement of the mean squared error (MSE), die mean absolute difference (MAD), the structural similarity index for images (SSIM), and more specifically the mean SSIM (MSSIM) [50]. It is well known that image difference measures such as taking the MSE do not correspond to the human perception of image difference [49]. Therefore, we employed the SSIM non-blind quality assessment. SSIM basically considers luminance, contrast, and structure terms to provide a measure of similarity for overlaid images. Figure 4.17 shows six pages of scanned images where independently acquired images of the same pages exist.

Results measured by MSE, MAD, and MSSIM are summarized in Table 4.4. All three measures are strictly larger for the SIFT/affine method than for FlexWarp. FlexWarp always achieves a MSSIM larger than 0.98, whereas the SIFT/affine method sometimes obtains significantly lower scores. The image *Codex Epist* gets a rather small MSSIM of 0.89677, which is due to the bending of the page, which is only adequately handled by FlexWarp. Figure 4.18 shows the SSIM image for registered images, where white indicates a low SSIM and black refers to a high SSIM. While white spots in the SSIM image for FlexWarp typically correspond to varying content caused by dust, etc., the SSIM for the SIFT/affine method clearly shows the page deformation as an artifact.

4.7   Conclusion

We described a real-time nonparametric registration method for printed matter in machine vision. The suggested FlexWarp algorithm automatically selects and assesses control points. A process similar to the well-known autocorrelation is used and called partial autocorrelation (PAC), where not the whole image is correlated but only the most salient part.

The main idea is that a proper CP exhibits low correlation values for all locations in his neighborhood, while an improper CP produces at least one high correlation value for a location in his neighborhood. We also studied the impact of distortions on the correlation values. It turned out that with comparable correlation values, bright image sections should be preferred for the normalized cross-correlation, while image sections with high contrast are preferable for the zero mean normalized cross-correlation.

We discussed image region size, search range, and the measure of similarity and specific applications in combination with CP selection. We provided individual limits for the CPs in order to indicate when a correlation value is so small that the CP should not be trusted. Our assessment functions are therefore well suited for automatic image registration.

The requirement of evenly distributing CPs over images was addressed. Our algorithm selects CPs from a candidate list. It does this iterative until the desired number of CPs is reached. A function that balances the requirements of high assessment values versus high distances to already selected CPs is used as selection criterion.

A practical real-time application of the online part of the algorithm in the field of print inspection was investigated. For this purpose, we implemented the time-critical parts of FlexWarp on a GPU and provided results for this parallel implementation.

We also addressed the integration and special requirements for a print inspection system in an industrial inspection system. Limiting factors for a typical inspection system with respect to maximum speed and requirements were discussed.

Finally, we provided implications on parameter settings and behavior of FlexWarp. We showed its superiority when compared to a general-purpose state-of-the-art automatic image registration approach.

References

1. Agrawal M, Konolige K, Blas MR (2008) CenSurE: center surround extremas for realtime feature detection and matching. In: Proceedings of European conference on computer vision. Springer, Berlin, pp 102–115
2. Aschwanden P, Guggenbühl W (1992) Experimental results from a comparative study on correlation-type registration algorithms. In: Förstner W, Ruweidel S (eds) Robust computer vision. Wichmann, pp. 268–289
3. Basler AG (2013) Ahrensburg, Germany: Basler sprint, user's manual for color cameras, 09 edn. URL: www.baslerweb.com
4. Bay H, Ess A, Tuytelaars T, Gool LV (2008) SURF: speeded up robust features. Comput Vis Image Understand (CVIU) 110(3):346–359
5. Bayer BE (1976) Color imaging array. US Patent 3,971,065
6. Beaudet P (1978) Rotational invariant image operators. In: Proceedings of international conference pattern recognition, pp. 579–583
7. Bodenstorfer E, Fürtler J, Brodersen J, Mayer KJ, Eckel C, Gravogl K, Nachtnebel H (2007) High-speed line-scan camera with digital time delay integration. In: Electronic imaging 2007. International Society for Optics and Photonics, p 64,960I
8. Bodenstorfer E, Hasani Y, Fürtler J, Brodersen J, Mayer KJ (2012) High-speed line-scan camera with multi-line CMOS color sensor. In: Computer vision and pattern recognition workshops (CVPRW), 2012 IEEE computer society conference on IEEE, pp 9–14
9. Bookstein F (1989) Principal warps: thin-plate splines and the decomposition of deformations. IEEE Trans Pattern Anal Mach Intell 11(6):567–585
10. Brown M, Szeliski R, Winder S (2005) Multi-image matching using multi-scale oriented patches. In: Proceedings of international conference on computer vision and pattern recognition, pp. 510–517

11. Burt PJ (1981) Fast filter transform for image processing. Comput Graph Image Process 16:20–51
12. Crow FC (1984) Summed-area tables for texture mapping. Comput Graph 18(3):207–212
13. Delaunay BN (1934) Sur la sphère vide. Bull Acad Sci USSR 7:793–800
14. Doytsher Y, Filin S, Ezra E (2001) Transformation of datasets in a linear-based map conflation framework. Surv Land Inf Syst 3:159–169
15. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24:381–395
16. Fitzgibbon AW (2003) Robust registration of 2D and 3D point sets. Image Vis Comput 21:1145–1153
17. Fonseca L, Kenney C (1999) Control point assessment for image registration. In: Proceedings of XII Brazilian symposium on computer graphics and image process, pp 125–132
18. Förstner W (1986) A feature based correspondence algorithm for image matching. Proc Int Arch photogrammetry ISP Comm 26(3):150–166
19. Gillman DW (1985) Triangulations for rubber-sheeting. Proc AUTOCARTO 7:191–199
20. Goshtasby AA (2012) Image registration. In: Similarity and dissimilarity measures. Springer, London, pp 7–66 (Advances in Computer Vision and Pattern Recognition)
21. Harris C, Stephens M (1988) A combined corner and edge detector. In: Proceedings of ALVEY vision conference, pp 147–152
22. Hartley RI, Zisserman A (2004) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, Cambridge. ISBN 978-0-5215-4051-3
23. He XF, Nixon O (2012) Time delay integration speeds up imaging. Photonics Spectra
24. Huber-Mörk R, Schindler A (2012) Quality assurance for document image collections in digital preservation. In: Proceedings of advanced concepts for intelligent vision systems (ACIVS), Brno, CZ, pp 108–119
25. Huber-Mörk R, Schindler A (2013) Automatic classification of defect page content in scanned document collections. In: Proceedings of international symposium on image and signal processing and analysis (ISPA). Trieste, IT
26. Kennedy R, Cohen W (2003) Automated designation of tie-points for image-to-image registration. Int J Remote Sens 24:3467–3490
27. Kersten T, Häring S (1998) Automatic tie point extraction using the oeepe/isprs test data—the swissphoto investigations. HafenCity Universität Hamburg, Hamburg (Report for the pilot center of the oeepe/isprs test, Photogrammetrie & Laserscanning)
28. Kitchen L, Rosenfeld A (1982) Gray-level corner detection. Pattern Recogn Lett 1(2):95–102
29. Krattenthaler W, Mayer K, Zeiller M (1994) Point correlation: a reduced-cost template matching technique. In: Proceedings of international conference on image process, vol 1. IEEE, NJ, pp 208–212
30. Krymski A, Tajima K (2006) CMOS image sensor with integrated 4gb/s camera link transmitter. In: IEEE international on solid-state circuits conference, ISSCC 2006. Digest of technical papers, pp 2040–2049
31. Laurini R (1998) Spatial multi-database topological continuity and indexing: a step towards seamless GIS data interoperability. Int J Geograph Inf Sci 12:373–402
32. Lepage G, Bogaerts J, Meynants G (2009) Time-delay-integration architectures in CMOS image sensors. IEEE Trans Electron Devices 56(11):2524–2533
33. Linde Y, Buzo A, Gray R (1980) An algorithm for vector quantizer design. IEEE Trans Commun 28:84–95
34. Liu XJ, Yang J, Shen HB (2008) Automatic image registration by local descriptors in remote sensing. Opt Eng 47(8):087, 206–(1–9)
35. Lloyd S (1982) Least squares quantization in pcm. IEEE Trans Inf Theory 28:127–135
36. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
37. Matas J, Chum O, Urban M, Pajdla T (2004) Robust wide baseline stereo from maximally stable extremal regions. Image Vis Comput 22(10):761–767

38. Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE Trans Pat Anal Mach Intell 27(10):1615–1630
39. Moravec H (1980) Obstacle avoidance and navigation in the real world by a seeing robot rover. Technical report CMU-RI-TR-3, Robotics Institute, Carnegie-Mellon University
40. Nosratinia A, Orchard MT (1996) Optimal warping prediction for video coding. In: Proceedings of international conference on acoustics, speech and signal processing, pp 1986–1989
41. Saalfeld S, Tomančák P (2008) Automatic landmark correspondence detection for ImageJ. In: Proceedings of imageJ user and developer conference, pp 128–133
42. Schaffalitzky F, Zisserman A (2002) Multi-view matching for unordered image sets, or "How do i organize my holiday snaps?". In: Proceedings of European conference on computer vision, vol I, pp. 414–431
43. Schmid C, Mohr R (1997) Local grayvalue invariants for image retrieval. IEEE Trans Pat Anal Mach Intell 19(5):530–535
44. Shi J, TomasiC (1994) Good features to track. In: Proceedings of conference on computer vision and pattern recognition, pp 593–600
45. Simoncelli EP, Adelson EH, Heeger DJ (1991) Probability distributions of optical flow. In: Proceedings of conference on computer vision and pattern recognition, pp 310–315
46. Sun W, Yang X (2008) Nonrigid image registration based on control point matching and shifting. Opt Eng 47(8):087,206–(1–9)
47. Teledyne Dalsa, Waterloo, Canada: Trilinear Camera PC-30-02K80-00-R, PC-30-02K60-00-R, PC-30-04K80-00-R, PC-30-04K60-00-R, Camera Users Manual (2011). URL: http://www.teledynedalsa.com
48. Tuytelaars T, Gool L (2004) Matching widely separated views based on affine invariant regions. Int J Comput Vis 59(1):61–85
49. Wang Z, Bovik A (2009) Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Process Mag 26(1):98–117
50. Wang Z, Bovik A, Sheikh H, Simoncelli E (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
51. Winter M, Fraundorfer F, Bischof H (2005) MSCC: maximally stable corner clusters. In: Proceedings of Scandinavian conference on image analysis, pp 45–54
52. Yang G, Stewart C, Sofka M, Tsai CL (2007) Registration of challenging image pairs: initialization, estimation, and decision. IEEE Trans Pattern Anal Mach Intell 29(11):1973–1989
53. Yu L, Zhang D, Holden EJ (2008) A fast and fully automatic registration approach based on point features for multi-source remote sensing images. Comput Geosci 34:838–848
54. Zitova B, Flusser J (2003) Image registration methods: a survey. Image Vis Comput 21 (11):977–1000
55. Zong J (2002) Multi-image tie-point detection applied to multi-angle imagery from misr. In: Proceedings of symposium on photogrammetry computer vision, ISPRS Comm. III, Graz, Austria, vol A, pp 424–429

Chapter 5
How Optical CMMs and 3D Scanning Will Revolutionize the 3D Metrology World

Jean-Francois Larue, Daniel Brown and Marc Viala

Abstract Every day, optical technology plays a bigger and bigger role in metrology. No doubt optical metrology will soon be used to perform continuous, high-speed inspections of items of all shapes, sizes, and materials. However, it will take time to move from a world dominated by coordinate measuring machines that use sensors to one where optical CMM reigns, especially in industries such as automotive or aeronautics where every new thing must prove itself first. One of the biggest changes in metrology in the past 30 years has been the development of portable measuring devices. This has brought inspection right into the production line, as close to the part as possible. The change—sparked by the development of portable measuring arms in the early 1990s and the emergence of laser trackers shortly after—turned traditional industries' inspection methods completely upside down. However, most portable measurement solutions still use technologies that have serious limitations in production environments. For example, these solutions demand extreme stability throughout the measurement process. Optical solutions, especially those using video cameras, sidestep these limitations by enabling automatic positioning and continuous measurement. Now, laser- and white-light-based digitizing technologies can also produce increasingly dense data. In this chapter, we will see the basic implementation principles and advantages offered by these optical solutions for performing direct inspections in a production environment. Key point is about using optical reflectors as metrological references on parts and using these references to (1) allow self-positioning of handheld scanners, (2) automatically align the device with a predetermined 3D reference, (3) easily move the device around the part, and (4) maintain high measurement precision in industrial production environments. We will also analyze all the factors that go into measurement precision and see how the use of optical technologies make it possible to greatly reduce the primary causes of measurement imprecision. To illustrate our analysis, the reader will find specific cases taken from real applications in the aeronautical, automotive, and naval industries.

J.-F. Larue (✉) · D. Brown · M. Viala
Creaform Inc, Québec City Area, 5825, rue St-Georges, Lévis, QC G6V 4L2, Canada
e-mail: jean-francois.larue@creaform3d.com

Contents

5.1   Introduction

Every day, optical technology plays a bigger and bigger role in metrology. For sure, optical metrology will soon be used to perform continuous, high-speed inspections for items of all shapes, sizes, and materials. Laser- and white-light-based digitizing technologies are producing increasingly dense and specific data, while the X-ray tomography-based technologies already in use can inspect the internal geometry of the most complex equipment.

However, from a realistic standpoint, it will take some time to move from a world dominated by coordinate measuring machines using tactile probes to one where optical CMMs reign, especially in industries such as auto manufacturing or aeronautics, where every new device must prove itself first. It is important to understand how sky-high are the stakes within these industries; the financial consequences of any production shutdown, delay, or defect are enormous, meaning that

change can only come about incrementally. This explains the success of solutions combining optics and tactile (laser trackers, optical trackers with sensors), which offer the benefit of optics (speed, portability, reduced sensitivity to the measurement environment and higher measurement volume) while remaining compatible with known and proven procedures. In the same vein, while digital solutions are being rapidly developed for working with uneven surfaces that sensors cannot easily accommodate, sensor solutions still remain widely used for inspecting geometric components in processes such as boring or stamping.

This essay will alternatively focus on two families of products: optical coordinate measuring machines (CMMs) and handheld scanners. We will begin by briefly recalling the evolution of both.

5.1.1 History of 3D Measurement Systems

5.1.1.1 Coordinate Measuring Machines (CMMs)

The first CMMs appeared in the 1960s and consisted mainly of 3D tracing devices equipped with a digital readout (DRO) used to display X, Y, and Z coordinates.

A number of companies claim to have invented the CMM. The first device was probably invented by the Italian company DEA (now part of the Hexagon Metrology Group), which introduced a CNC portal frame CMM equipped with a probe and DRO in the late 1950s.

Today, myriad different types of CMMs are available, including cantilever CMMs, bridge CMMs, gantry CMMs, horizontal arm CMMs, portal CMMs, moving table CMMs, fixed bridge CMMs, and articulated arm CMMs. Over 6,000 CMMs are still produced every year. There are two main categories of CMMs: manually operated models where an operator guides the probe in contact with the part and models that are guided automatically by a program [1].

CMM measuring arms are equipped with either fixed probe heads (on manual CMM models, operators guide probe heads manually) or touch-trigger probe heads (which begins measuring automatically when the probe comes into contact with the part). Automated CMM technology was invented in the 1960s to be used for complex inspections of Concorde supersonic jet engines. This invention led to the creation of the company Renishaw in 1973, now the main supplier of CMM measuring heads [2].

Inspection software is needed in order to use a CMM, and CMM manufacturers originally provided the software along with the machine. Today, numerous types of inspection software are available from independent companies not affiliated with the machine manufacturers. They are compatible with most CMM brands and have enabled big companies with CMM fleets to standardize operation of their machines, reducing operator-related training costs considerably. Standardization also makes it possible to replace older machines with more recent models or newer technology without losing the many existing inspection programs or requiring operators to

undergo lengthy training. The main software applications are all capable of inspecting prismatic and geometric entities in addition to lines or surfaces.

Operating a CMM requires a qualified technician and a controlled environment (vibrations, temperature, dust), and CMM accuracy inspections must typically be conducted every 12 months. Many standards have been established to regulate CMM accuracy, including B89 in the USA, VDI in Germany, and of course the ISO 10360 international standard.

Today, CMMs are most commonly used for measuring manufactured parts in 3D, reverse engineering, and digitizing prototype parts. However, conventional CMMs can be used only in very specific environments like a metrology laboratory or similar facility where temperature is very precisely regulated and vibrations are filtered using a vibration isolation pad.

The introduction of portable CMMs with measuring arms in the 1980s marked a new revolution in measurement: Finally, it was possible to bring 3D measuring to the production line. This innovation eliminated the need to move manufactured parts to a dedicated, controlled environment, making this type of CMM very popular. However, because the measuring arm still used a classic technology based solely on precision mechanical parts, it was fragile. These types of CMMs remain very sensitive to the vibration environment, and considerable precautions are required when operating them. That is why the twenty-first century is destined to be the age of optical technologies. Optical solutions offer excellent portability, faster measurements, and superior reliability at an extremely competitive price (Fig. 5.1).

For a long time, optical three-dimensional measuring processes, which were limited to the use of theodolites, tacheometers, and conventional photogrammetry rooms (gelatin silver process), were almost exclusively used in classic geodesy and cartography.

The main reason behind this "limited" field of application had to do with the methods involved: Optical 3D did not deliver immediate measurement results because the data recorded by the sensors could not be directly transcribed into the desired *XYZ* coordinate points. Most sensors provided angles that had to be translated through geometric calculations into Cartesian coordinates, taking into account a certain amount of data (triangulation, angle composition and distance,

Fig. 5.1 CMM evolution

transfer of the sensor referential to the object referential, and other calculations). Consequently, and until such time computer tools became easily transportable, these methods remained "reserved" for fields where no other measuring methods were available.

In the mid-1970s, with the rapid development of computer technology, industry began turning to photogrammetry as it became possible to mathematically correct for the distortions in the optics used and perform complex calculations on desktop computers. But photogrammetry remained marginal because of the still significant delay between measurement taking and results, and also because industry was skeptical of these unorthodox methods.

It was not until the mid-1980s that a series of new developments emerged to facilitate industry adoption of these new measuring methods, notably the arrival of the first electronic theodolites, the use of scanners to scan gelatin silver plates, the advent of the PC as the primary computer tool to accelerate photogrammetry data processing, the subsequent arrival of the first electronic distance meters, and the miniaturization of video cameras. Industries where these methods were most invaluable (aerospace, automotive, shipbuilding, and nuclear plant maintenance, for example) quickly integrated them into their production processes, despite a few initial reserves and obstacles.

The main obstacle then was the difficulty most users had in mastering measurement uncertainties, which, in a field of extremely stringent quality standards, could be a deal breaker. These methods have now reached maturity, and a certain number of studies and analyses have been conducted, but the obvious lack of normative documents still sometimes makes it difficult to implement these methods broadly in these industries [3].

5.1.1.2   Handheld Scanners

The mechanical and optical solutions discussed thus far can only be used to measure 3D points, most often through contact with a probe, and sometimes by marking the point with a retroreflector or an LED. This process is not suited for surface inspections, because it is too slow when a large number of points are required, particularly for skew surfaces (the fastest CMMs can measure hundreds of points per second).

The National Research Council of Canada was one of the first institutes to develop laser scanners in 1978 [4].

In the early 1980s, 3D scanning solutions appeared that were capable of creating files of points, or "point clouds" (which are later processed into computer mesh data, such as STL files) based on the measured surface.

In the mid-2000s, a major breakthrough took place in the laser scanner industry, with the appearance of self-positioning laser scanning systems. These three-dimensional devices use no external positioning systems. Instead, they use retroreflective targets placed on the object as a positioning system.

5.1.2   Recent Advances

Over the past 30 years, one of the biggest changes in metrology has been the development of portable measuring devices. This has brought inspection right into the production line, as close to the part as possible. The change—sparked by the development of portable measuring arms in the early 1990s and the emergence of laser trackers shortly after—turned conventional industry inspection methods completely upside down. It also made it possible to take measurements more quickly and more often, fostering huge improvements in response time and quality.

Over the past decade, the pace of change has been accelerated even more by the development of the arm-mounted scanners and self-positioning handheld 3D scanners that now dominate the market because of their portable digitization capabilities.

However, most portable measurement solutions still use mechanical technologies that present serious limitations in production environments. For example, these solutions require extreme stability throughout the measurement process, which means they require the use of costly, cumbersome, and generally inflexible control templates and heavy bases—including the granite tables long associated with CMMs.

Optical solutions, especially the ones using video cameras (technologies evolving out of photogrammetry/videogrammetry), sidestep these limitations by enabling automatic positioning and continuous device measurement (the self-positioning concept or dynamic referencing). While they are generally available with most optical technologies, there are some restrictions, as in the case of laser trackers.

Self-positioning handheld scanners are the perfect example of this technology, since they make it possible to completely forget external means of positioning.

5.1.3   Current Industrial Context

It is interesting to note that scientific and technological advances have always been tightly coupled with the improvement of measuring systems. The twenty-first century has already witnessed the explosion of the ability to measure in several domains. Probably, the most significant examples are the experiments with the CERN's Large Hadron Collider, each generating several terabytes of observation data.

Closer to our domain, existing 3D optical sensing technologies allow quick generation of a tremendous amount of dense points on the surface of an object. Indeed, capturing 20,000 to more than a million 3D points per second is no longer exotic. This ever-growing capability to collect observations combined with the improved computational technologies not only paves the way to, but inevitably leads us toward the new paradigm of intelligent measurement.

5.2 Photogrammetry Overview

5.2.1 Background

3D photogrammetric calculations make it possible to measure an object in three dimensions using a set of photographs taken from different viewing angles. This technique, known as bundle adjustment, was used in optical measuring applications in the early 1960s. An extensive literature on photogrammetry [5, 6] describes this process of numerical computation based on nonlinear optimization; more recently, the technique was revisited in the field of computer vision [7, 8], and numerous applications have appeared for dense reconstruction using a sequence of images [9, 10], for visual localization or odometry [11, 12].

Figure 5.2 illustrates the bundle adjustment principle where four photographs of a scene are used to perform a joint estimate of the coordinates of the geometric entities in the scene, here the coordinates of the points (X_0, Y_0, Z_0, …, X_3, Y_3, Z_3) and camera poses (P_0, P_1, P_2, P_3). In most applications, the parameters of the model describing the image acquisition device are also considered as variables that must be identified (i.e., auto-calibration process).

Unlike other optical measuring techniques, photogrammetric devices do not make it possible to directly infer the scale of the scene: This means the point coordinates and camera poses will be estimated up to a scale factor $\lambda \neq 0$. In order to establish the scale of measurement, one dimension of the scene must already be
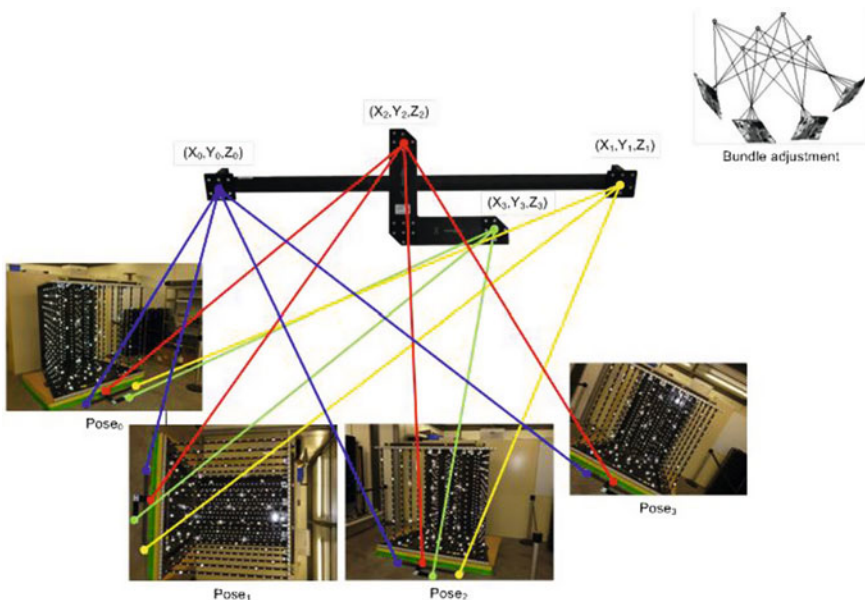
Fig. 5.2 Bundle adjustment principle

known or a length standard must be placed in the scene and be visible in the sequence of images. However, one of the specific characteristics of photogrammetry is its ability to effectively measure medium- or large-sized objects (from a few meters to a few hundred meters in size) using a length standard and an adapted acquisition sequence.

5.2.2 Photogrammetric Principle and Calculation Modules

The final photogrammetric calculation (i.e., bundle adjustment) requires the following:

- Detection of targets in the measurement session images (detection refers to the target pixel coordinates, which must be provided by the image-processing module);
- Pairing of targets in the measurement session images (each detection corresponds to a target in the scene, and the role of the association module is to pair each detection with its corresponding target for each image); and
- Approximated camera poses and target coordinates of the measurement session. Bundle adjustment is carried out via nonlinear minimization, which requires a good initialization of the variables to be estimated. An iterative reconstruction module must be developed to perform these calculations and generate the best partial model of the scene for each acquisition.

Figure 5.3 illustrates the various calculation steps required for a photogrammetric engine. Moreover, in order to guarantee maximal automation of the solution,
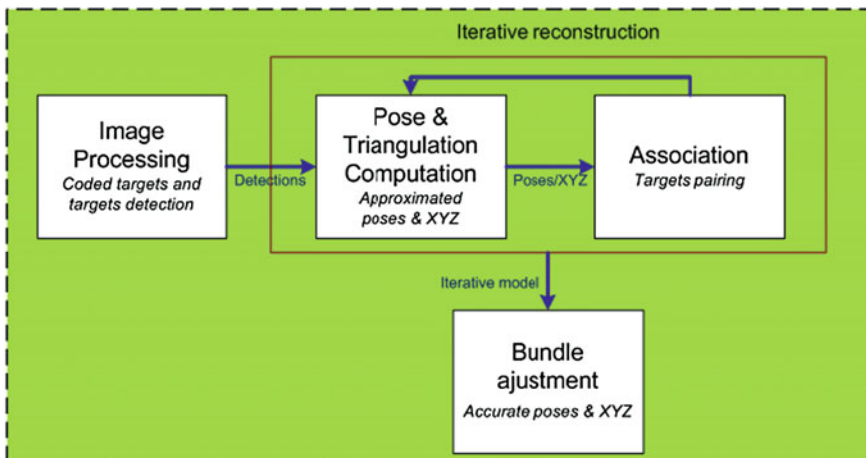
Fig. 5.3 Required calculation for a photogrammetric engine

the photogrammetric engine must make full use of coded targets, because they have properties that are important and useful in this process:

- Coded targets in an image are robustly identifiable by their unique signatures. With hundreds of different codes available, the image-pairing phase is thus trivial and immediate.
- In theory, if we have the coordinates of one coded target—itself made up of 7 single targets in our case—we can reconstruct or calculate the pose. In practice, however, it is best to have at least three coded targets per image. This property is important because it allows us to design a reference artifact with coded targets to initialize the photogrammetric reconstruction process. Once this process is initiated, we need to have at least one coded target per image to be able to measure and reconstruct it before we begin pairing its single targets.

Next, we consider the intrinsic parameters of the model associated with the image-recording device are estimated during a calibration step. This step is completed during the production of the device and is similar to the camera calibration step we use for our HandyPROBE product line. We will look at the contribution of these procedures later. A prior knowledge of these intrinsic parameters is important and vital for ensuring the success of the numerical calculations involved by the bundle adjustment process. From this perspective, our implementation method is very similar to that used by photogrammetrists. During bundle adjustment, all parameters will be optimized, including the model of the image-recording device, in order to avoid any optomechanical drift that may be introduced during use and transportation.

5.2.3 Bundle Adjustment

5.2.3.1 Principles

The principle of bundle adjustment consists of accurately measuring a scene structure using a sequence of images acquired by a camera or a series of cameras. This calculation method was presented and developed in the 1970s, but more recent research is available [13]. Studies have shown that bundle adjustment remains a highly effective and optimal approach.

It is based on a simultaneous estimation of the scene's structural parameters in the least-squares sense—e.g., the scene point coordinates—and the acquisition parameters, i.e., the poses and parameters specific to the camera(s), known as extrinsic and intrinsic parameters, respectively. For our purposes, we will focus only on measuring coordinates of points materialized by retro-targets.

Since the problems we face can grow considerably, adapted procedures have been studied and implemented.

Recent studies have also looked at ways to deal with even bigger problems involving thousands of images with a scene containing tens of thousands of points [14].

5.2.3.2 Nonlinear Optimization—Levenberg-Marquardt

Let us consider a measurement situation involving a scene made up of N points whose coordinates are given by vectors X_p, $p = 1$, ..., N, and are perceived in M images, with each image characterized by one camera pose (orientation and position) P_i, $i = 1$, ..., M, and camera model K_c, $c = 1$, ..., L. In this context, the camera model parameters are considered constant throughout image acquisition.[1] The camera model incorporates perspective projection and additional parameters to get rid of optical aberrations caused by the lens (radial distortion, tangential distortion, etc.). Each point perceived in a given image yields an uncertain observation or measurement x_{ip} and a theoretical projection provided by the prediction model $\hat{x}_{ip} = \mathbf{f}(K_c, P_i, X_p)$. We can then express the prediction error between the measurement and its predictive model:

$$\Delta x_{ip}(K_c, P_i, X_p) = x_{ip} - \hat{x}_{ip} = x_{ip} - \mathbf{f}(C_c, P_i, X_p) \tag{5.1}$$

The principle of bundle adjustment is to minimize prediction error, or residual error, in the least-squares sense by considering all points and their measurements; the variables of this minimization are the parameters to be estimated (i.e., the point coordinates, poses, and intrinsic parameters) $\mathbf{x} = (K_c, P_i, X_p)$. Then, we consider that the prediction errors will be represented by vector $\mathbf{e}$. By adopting the Levenberg-Marquardt method, this minimization boils down to an iterative process where each iteration consists of calculating a correction $\delta \mathbf{x}$ of vector $\mathbf{x}$:

$$(\mathbf{J}^t\mathbf{J} + \lambda \mathbf{l})\delta \mathbf{x} = -\mathbf{J}^t\mathbf{e} \tag{5.2}$$

where $\mathbf{J}$ represents the Jacobian matrix of the prediction error

$$\mathbf{J} = \frac{d\mathbf{e}}{d\mathbf{x}} = -\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}. \tag{5.3}$$

At the convergence, when the corrections become lower than a preset threshold, the estimated parameters are given by vector and their a posteriori uncertainty is modeled by the covariance matrix, which is approximated as follows:

$$\mathbf{Cov}(\mathbf{x}) = (\mathbf{J}^t\mathbf{J})^{-1} \tag{5.4}$$

---

[1]This hypothesis requires the use of an optomechanical stable acquisition device.

In a joint estimate of points, poses, and cameras, the estimation problem suffers from an ambiguity in scale and pose, or 7 degrees of freedom, [6, 8, 15]. This ambiguity means that $\mathbf{J}^t\mathbf{J}$ is not a full rank matrix and it cannot be inverted without either

- calculating its pseudoinverse;
- taking into account the principle of gauge freedoms and internal constraints introduced by photogrammetrists[2]; and
- using a regularization method that removes this deficiency.

The Levenberg-Marquardt method belongs to the last category. By "augmenting" normal equations, it adds a positive perturbation, λ, to the diagonal elements of $\mathbf{J}^t\mathbf{J}$, cf. (5.2). This frees us from having to calculate the pseudoinverse, a numerically complex operation, and allows us to implement a classic linear system resolution method. For matrix $\mathbf{J}^t\mathbf{J}$, two "augmentation" methods are possible:

1. The first consists of adding the same constant term to all diagonal elements (see Footnote 1). This is the approach used in the Levenberg-Marquardt article. It guarantees the proper conditioning of "augmented" matrices, ensuring that matrices will always be full rank.
2. The second approach consists of adding to the diagonal elements a term proportional to their current values $(\mathbf{J}^t\mathbf{J} + \lambda\,\mathrm{diag}(\mathbf{J}^t\mathbf{J}))$. This approach may be useful but presents the disadvantage of not guaranteeing the non-singularity of the "augmented" matrices [7].

5.2.4 MaxSHOT 3D/VXSHOT Photogrammetric Solution

The MaxSHOT 3D/VXshot photogrammetric solution is composed of VXshot software and MaxSHOT 3D image acquisition equipment. Other artifacts were also required to implement the photogrammetric solution: reference object, coded targets, and a scale bar. All of the equipment is presented in Fig. 5.4.

The artifacts were developed to allow for full automation of the photogrammetric measuring process:

- an L-shaped artifact allowing us to make approximate calculations of poses if it is visible or partially visible;
- magnetized coded targets to ensure all images are oriented; and
- a magnetized scale bar equipped with two coded targets to ensure its complete identification (its center-to-center length is calibrated).

Numerous tests were conducted on scenes of varying degrees of complexity (see Fig. 5.5).

---

[2]This principle of resolution widely used by the photogrammetry community requires the addition of constraint equations to the system to be resolved, making its numerical resolution more complex than it was initially.

MaxSHOT 3D

Scale Bar

Set of coded targets

L-shaped artifact

Fig. 5.4 Equipment used in the MaxSHOT/VXshot photogrammetric solution

VDI artifact
1.97m (L) × 1.55m (H) × 1.97m (D)
49 reference lengths to measure
274 coded targets and 540 targets

Long object
2.9m × 17m
16 reference points measured with a
laser tracker for a total of 120 reference
lengths
217 coded targets

Fig. 5.5 A few reference sessions to test the photogrammetric solution

5.3 Optical CMMs

5.3.1 Overview

Optical CMMs using image-based triangulation are constituted of an optical tracking system equipped with video cameras. There are two categories of optical CMMs: those based on matrix array cameras (which use retroreflective targets or LEDS as targets) and those with linear array cameras (which systematically use LEDs as targets). This paper deals only with the first category.

Optical CMMs use three main calculation steps to implement their dynamic referencing and optical measurement functionalities:

An image-processing step to accurately assess target image projections in the stereo sensor images;

A triangulation step to estimate the target coordinates (X,Y,Z) in the sensor reference from their projections in the twin images of the stereo sensor; and

A step to estimate the pose of a modeled object using a set of points whose nominal coordinates are known and observed by stereo sensor.

The cameras can be modelized using perspective projections and other parameters that take into account geometric aberrations generated by the imager (camera with its lens), commonly known as radial and tangential distortions. Identification of these so-called intrinsic parameters occurs during a previous calibration step. The approach used to calibrate the imagers is based on techniques used in photogrammetry [6].

With this calibration, the projection of a 3D point $\mathcal{P}$ expressed in the stereo sensor reference in one of the imagers is given by the following perspective projection relationship [7, 16]:

$$\mathbf{x} = \boldsymbol{P}\mathbf{X} \tag{5.5}$$

where $\mathbf{X} = (X, Y, Z, 1)^{\mathrm{t}}$ represents the homogenous coordinates of $\boldsymbol{P}$; $\mathbf{x} = (u, v, 1)^{\mathrm{t}}$ the coordinates of the projection; and $\boldsymbol{P}$ the projection matrix:

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_1 \\ \boldsymbol{P}_2 \\ \boldsymbol{P}_3 \end{pmatrix} = \boldsymbol{K} \times \left(\begin{array}{c|c} \boldsymbol{R} & \boldsymbol{t} \\ 0_{1\times3} & 1 \end{array} \right) \tag{5.6}$$

$\boldsymbol{K}$ being the camera matrix integrating only its intrinsic parameters (we ignore here the additional radial and tangential distortion correction parameters):

$$\boldsymbol{K} = \begin{pmatrix} A_u & 0 & C_u \\ 0 & A_v & C_v \\ 0 & 0 & 1 \end{pmatrix} \tag{5.7}$$

And where $\boldsymbol{R}$ and $\boldsymbol{t}$ represent the rigid transformation (rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$) ensuring the transfer from the sensor reference frame to the imager reference frame.

This relationship is the basis for calculating the coordinates of point $\boldsymbol{P}$ by triangulation and for calculating the pose of an object.

The image-processing step consists of extracting the images acquired by the stereo sensor, i.e., the target center projections observed in the sensor environment. This step significantly influences optical sensor performance, both in terms of repeatability and accuracy.

Several approaches are documented in the literature [17–19]. Some involve calculating the grayscale barycenter of the target light response, while others involve approximating the response curves by ellipses or modelizing the responses. In the latter cases, estimation of the model is based on a least-squares estimator between the pixel value in image I and the pixel value determined by the model $\hat{I}(\boldsymbol{p}_{\mathrm{model}})$:

$$\mathcal{C}_{\min}(\boldsymbol{p}_{\mathrm{model}}) = \sum(\hat{I}(\boldsymbol{p}_{\mathrm{model}}) - \boldsymbol{I})^2 \tag{5.8}$$

The photogrammetric step after sub-pixel target detection is triangulation; in this step, the traces of a target captured in the two stereo sensor images are measured and used to calculate the target coordinates (X, Y, Z) in the sensor reference.

Numerous techniques are available to infer coordinates of a target from its projections in a pair of stereo images [6, 7]. The simplest is to calculate the midpoint. Geometrically, this solution consists of taking the midpoint of the common perpendicular joining the two optical rays running through the traces and the target (see Fig. 5.6).
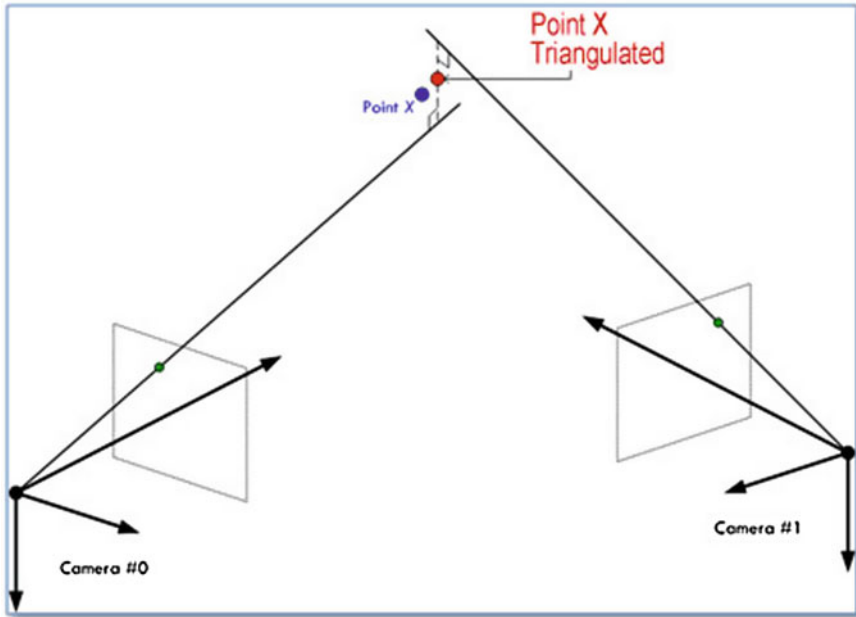
Fig. 5.6 Principle of midpoint triangulation

Other approaches involve solving a linear system using perspective projection relations (5.1) or (5.2) solving a polynomial equation [20].

Once target triangulation is complete, targets are identified as belonging to a rigid body or not: A rigid body can be described as a series of targets rigidly linked in their own reference frame. Identification is followed by 3D localization of the rigid bodies through measurement of their 3D targets and nominal coordinates; 3D localization makes it possible to estimate the position and orientation of the rigid body's reference frame, or pose, in relation to a measurement reference, in this case the sensor reference. Once 3D localization of the rigid body has been estimated, the position of the probe center is deducted.

The 3D localization step is crucial to the measurement performance and can be based on a linear localization method that models rigid body orientation using a quaternion [21]. This method has been shown to be effective in simulations [22].

The other possible technique tackles a 2D/3D calculation problem by drawing not on the coordinates of measured targets, but rather on the targets' projections in the stereo sensor imagers. This technique, known as resection, is widely used in photogrammetry [5] and has the advantage of minimizing error directly in the target measurement space. In the computer vision literature, this problem is known as the PnP[3] problem. In its initial formulation, it is nonlinear if we consider the

---

[3]For Perspective N-Point.

perspective projection relationship (see Footnote 3), and can be solved by minimizing a criterion in the least-squares sense:

$$C_{\min}(\widehat{\boldsymbol{R}},\widehat{\boldsymbol{t}}) = \sum_{i=0}^{N} \left\| \begin{pmatrix} {}^{0}\mathbf{f}(X_j) - {}^{0}\boldsymbol{x}_i \\ {}^{1}\mathbf{f}(X_j) - {}^{1}\boldsymbol{x}_i \end{pmatrix} \right\|^2 \tag{5.9}$$

where $(\widehat{\boldsymbol{R}},\widehat{\boldsymbol{t}})$ represents the pose rotation and translation parameters to be estimated, $(X_1, \ldots, X_N)$ the nominal coordinates of N points in the object reference, and $({}^{0}\boldsymbol{x}_1, \ldots, {}^{0}\boldsymbol{x}_N)$ and $({}^{1}\boldsymbol{x}_1, \ldots, {}^{1}\boldsymbol{x}_N)$ the coordinates of their projections in the first imager and stereo sensor, respectively. ${}^{j}\mathbf{f}(\ldots)$ represents the perspective projection function for j given by Eqs. (5.5) and (5.6), that is to say:

$$ {}^{j}\mathbf{f}(X_i) = \begin{pmatrix} \frac{{}^{j}\boldsymbol{P}_1(X_i)}{{}^{j}\boldsymbol{P}_3(X_i)} \\ \frac{{}^{j}\boldsymbol{P}_2(X_i)}{{}^{j}\boldsymbol{P}_3(X_i)} \end{pmatrix} \tag{5.10}$$

The rigid body concept applies in equal measure to probes and scanners that are identically instrumented with targets and localized in 3D by an optical localization system.

Once the calculation is performed, we can deduce the position of the center of the probe ball at all times (using a prior calibration step to determine the position of the ball center in the reference defined by the probe targets). The same deduction can be made if the tip is replaced by an optical probe using a laser line projection and camera (scanner) (Fig. 5.7).
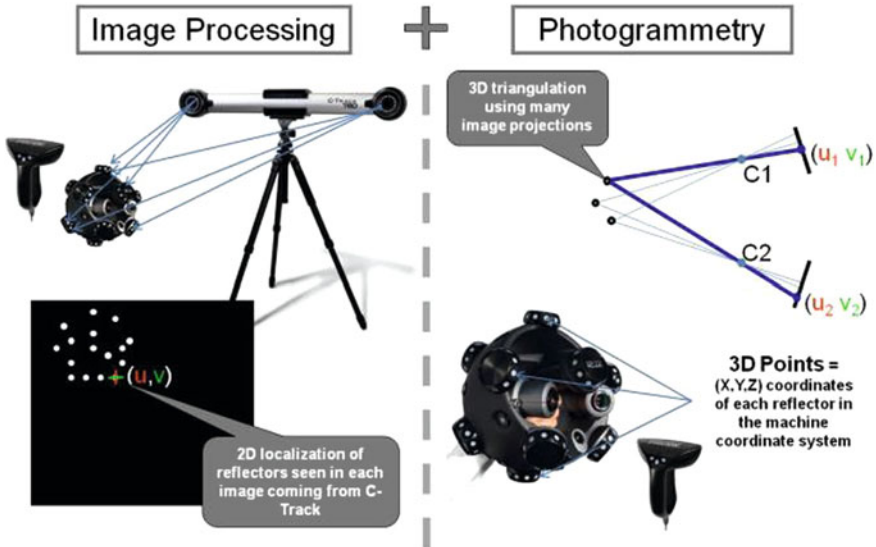
Fig. 5.7 Operating principles of an optical CMM using image-based triangulation

Cameras in an optical localization system can simultaneously observe other targets than those present on the probe. For example, targets can be positioned in advance on the object to be measured. Since these targets are measured at the same time as the probe targets, it is possible to calculate not only the probe position, but also the part position at the time a point is probed. Since matrix cameras are used, all targets are also triangulated synchronously at a frequency of 30 Hz (whereas certain optical localizers use linear cameras that can only acquire targets sequentially). This makes it possible to calculate probe position in direct relation to the part, and not just the machine reference. All measuring operations are thus performed directly in the part reference and will not be affected by possible movements —even rapid ones—of either the part or the optical localization system. This functionality is known as dynamic referencing. It is the core of the TRUaccuracy technology featured in Creaform products.

Thanks to this technology, optical CMMs can be mounted on a light tripod for unmatched portability (in the case of handheld, self-positioning scanners, the tripod is dispensed with altogether). The operator can thus move freely around the part for easier measurement of hard-to-access points, without any loss of alignment.

Most of all, in a shop environment affected by extensive vibrations, as noted above, dynamic referencing makes it possible to maintain levels of uncertainty normally found only in metrology laboratories. As we shall show below, as a result of dynamic referencing, vibrations have no negative effect on measurement uncertainty.

5.3.2 *Dynamic Referencing*

While conventional CMMs measure reference points only once at the beginning of the measuring phase, optical CMMs measure optical reflectors continuously and simultaneously with each point (which is equivalent to realigning the part as each point is measured) (Fig. 5.8).

Measurements are thus taken as if the part were attached to the optical CMM—in exactly the same way it would if the part were attached to the table of a conventional CMM.

Instead of optical reflectors, it is also possible to use simple adhesive retroreflective targets, which can be placed anywhere on the part. However, optical reflectors do present a number of advantages over these targets:

- Since the reflector is perfectly centered on the axis of the optical reflector, it is possible to move the CMM around the part without losing alignment. Simply turn the reflectors on their axes so they remain visible to the CMM.
- Optical reflectors can be used in place of mechanical reflectors for precise position locating (on a tool, for example). This comes in very handy when it comes to automatically aligning the part or the tool, as we will see in the next section.

Fig. 5.8 Principles of dynamic referencing

5.3.3 Automatic Alignment

We will once again refer to the optical reflectors described above, but this time, we will place them in precisely defined locations that have been measured beforehand. That means we know the 3D coordinates of these locations in the optical reflectors referential (measuring referential, CAD referential). Providing these coordinates to the CMM will effectively link the CMM work referential to the part referential for the entire duration of the measurement (thanks to its dynamic alignment functions). From that point on, all the points measured will be expressed directly in the part referential, eliminating the need to measure points by probing to align the part. This not only saves a great deal of time, but also and more importantly, gets rid of all operator-based alignment errors on alignment points. These kinds of operator errors are very common and result in inspection defects that are often difficult to spot.

5.3.4 User Calibration

Another benefit brought by some certain optical CMMs is the ability to quickly calibrate the CMM directly at the work site. With the HandyPROBE™ CMM, calibration only takes minutes using a calibration bar, itself certified annually by a laboratory, which is a member of the International Committee for Weights and Measures. Operators position the calibration bar, which is equipped with several reflectors at known distances, at various locations within the CMM measurement volume. The CMM automatically identifies and uses the resulting data to immediately compensate for any measurement volume errors. This provides the operator

with guaranteed stability and accuracy for the entire duration of use, whereas with other annually calibrated portable CMMs, highly restrictive shop floor environments can cause difficult to identify drifts.

5.3.5   Increasing Measurement Accuracy and Decreasing Causes of Imprecision

Measurements can be affected by a number of factors, including:

- The measuring device,
- The measuring environment (temperature, vibrations, etc.),
- The operator (skills, stress level, etc.),
- The measuring method, and
- The part being measured (Fig. 5.9).

An easy way to remember this is to think about the 5M (1-Machine, 2-Man, 3-Measured part, 4-Method, and 5-Measuring conditions). Different device types can be linked to different factors; for instance, conventional CMMs, measuring arms, and laser trackers are very sensitive to the environment (measuring conditions), whereas laser trackers require highly qualified operators, and measuring arms sometimes need leapfrog methods to measure larger parts, which can be a major source of errors.

Operator-related errors are a particularly important factor in poor quality measurements. This is in fact one of the main conclusions of the 2011 CMSC

Fig. 5.9 The 5M law
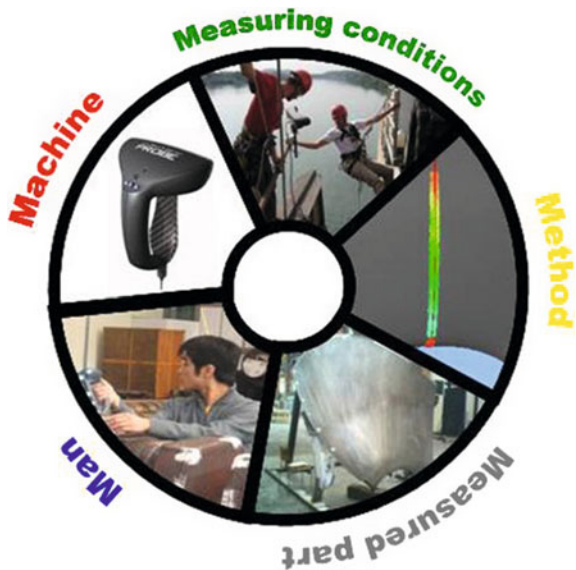
Measurement Study Report, "How Behavior Impacts Your Measurements," a compelling and detailed analysis of operator behavior in the metrology process.

More specifically, the study indicates that only 6 % of participants in the "Day 1: Articulated Arm/Engine Compartment" test noted that the measurement device was on a carpet.

In the "Day 1: Laser Tracker/Door" test, only 6 % paid attention to the stability of the part. In this same test, analysis showed that 7 % of participants moved the part after aligning and that only 7 % checked drift on one alignment point at the end of measurement.

In the "Day 1: Laser Tracker/Vehicle" test, 20 % expressed concerns about the presence of the carpet, 15 % questioned part stability, and only 25 % mentioned the need to obtain a good alignment.

On the first day of testing, no instructions or procedures were given to participants. Yet more than 40 % of the participants worked in the quality control or inspection field, and over 60 % had at least 7 years of experience and/or performed measurements on a daily or weekly basis.

On the second day, instructions and procedures were provided to the participants. The impact was positive, because 90 % of those who took the "Day 2: Articulated Arm/Engine Compartment" test identified the problems posed by the carpet and the large number of people in the measurement area. Nonetheless, less than 30 % expressed concern about the stability of the part.

In the Day 2: Laser Tracker/Door test, over 80 % of the participants identified the problem with the carpet, but less than 40 % expressed concern over the large number of people in the measurement area.

Overall, though, the instructions about measurement quality had a very significant effect, with notable improvements in the deviations observed on day 1 and day 2:

Engine compartment: Day 1: 0.56–3.81 mm and
                     Day 2: 0.15–0.90 mm;
Door: Day 1: 35–43.18 mm and
                     Day 2: 0.13–0.24 mm;
Vehicle: Day 1: 5.36–8.198 mm and
                     Day 2: 0.926–1.59 mm.

It is impossible to determine the contribution of errors related to the instability of the measurement setup, but they probably contributed significantly to the errors observed on day 1.

Dynamic referencing actively contributes to reducing some of the human errors identified in the CMSC study, i.e., inadequate operator assessment of the risks that derive from an unstable environment, heavy traffic, or an unstable part.

So when deciding which measuring solution to use, it is important to not only take the accuracy of the device into account, but also its ease of use for a production operator and its overall performance under factory measuring conditions that are more restrictive than a laboratory.

Triangulation-based optical CMMs are very easy to use because, unlike measuring arms, they have no mechanical link. And unlike laser trackers, which require the implementation of a specific procedure to enable the tracker to lock back in on the prism when the laser beam is cut off by an obstacle, this type of CMM is tolerant of beam cutoff. With an optical tracker, even if the probe is accidentally obscured by an obstacle, it will be automatically detected once it is visible again.

But the biggest advantage of this solution is its robustness in typical shop floor environments—yielding a degree of measurement precision equivalent to that achieved in the laboratory using a conventional solution.

5.4 Handheld Scanners

5.4.1 Overview

Handheld scanners rely on two cameras that, in addition to triangulating the laser line, are also used for triangulating the object being measured—more specifically to determine the scanner position in relation to specific points placed on the object (i.e., retroreflective targets) (Fig. 5.10).

Certain scanners such as the Handyscan 3D laser scanner from Creaform even generate an STL file directly. This makes it possible to quickly generate a dense representation of shape imperfections (the fastest scanners can acquire hundreds of thousands of points per second), which has considerably advanced our understanding of certain fabrication processes (drawing, smelting, molding, etc.). Nowadays, there are two main categories of scanners: structured light scanners and laser scanners. These will be discussed in the next section.

Fig. 5.10 Self-positioning laser scanner measuring a casting part
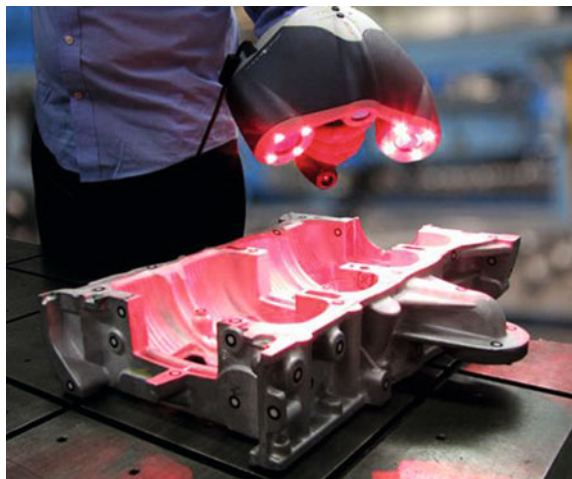
5.4.2 Structured Light Scanners

Structured light 3D scanners project a pattern of light onto a part and observe how the pattern is distorted. The light pattern is projected by either an LCD projector or a scanned or diffracted laser beam. One or two cameras are used to observe the projected pattern.

If only one camera is used, the position of the projector in relation to the camera must be determined in advance; if two cameras are used, it is enough to calibrate the stereoscopic pair.

Speed is the biggest advantage of structured light, because it is possible to acquire thousands of points in a single acquisition rather than having to scan the points sequentially. Structured light also makes it possible to acquire unstable or moving objects, or even track distortions [23] (Figs. 5.11 and 5.12).

As with optical scanning methods, reflective and transparent materials are problematic. This obstacle is most often overcome by placing a powder on the material in question to yield a matte surface [24].

5.4.3 Laser Scanners

Laser scanners project a point or a line on the object, but in this case, the pattern is a simple line. The point or line is observed by one or two cameras, making it possible
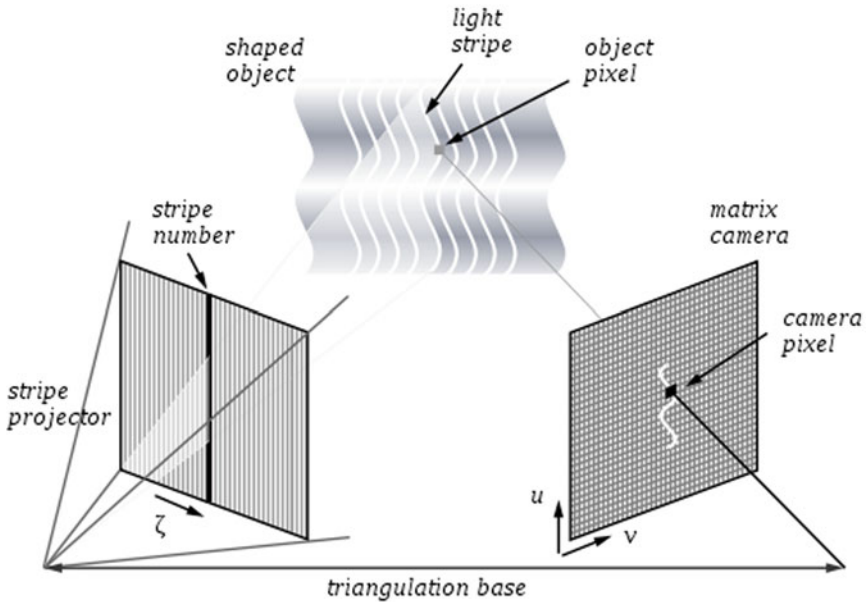
Fig. 5.11 Fringe scanner basics (from Wikipedia)

Fig. 5.12 Example of fringe projection on a part (from Wikipedia)
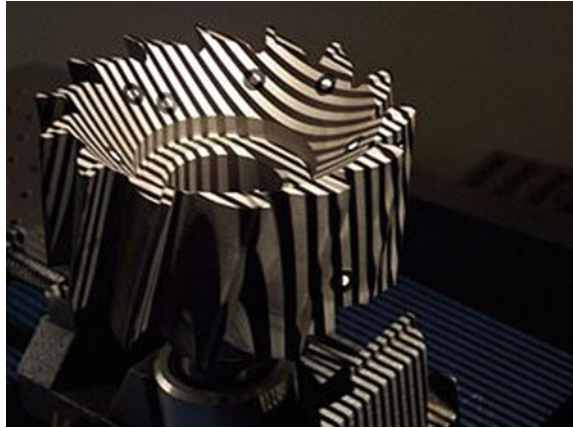
Fig. 5.13 Distance estimation using the geometry of the pair camera—laser
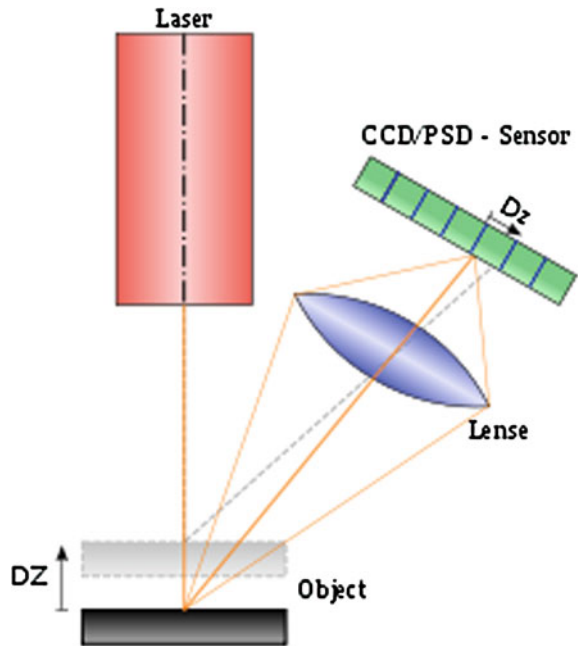
to triangulate the 3D coordinates of the projected point. The surface is reconstructed progressively by moving the scanner in order to sweep the surface of the object with the projected line [25] (Fig. 5.13).

5.4.4   Triangulation Basics

If we look at the evolution of three-dimensional measurement, we observe an improving trend in the speed and density of measurements. In a number of applications (measuring complex or free-form parts, for instance), the measurements needed to conduct the analysis or test must be information dense, but should not unduly lengthen the required measuring time. This is particularly true when measuring parts as part of a first-article inspection or an in-depth tooling control process, because these types of metrology situations require the collection of dense data on all areas of the part.

This need is what sparked the invention of triangulation-based laser scanners, which make it possible to shift from point-by-point three-dimensional measurements to fast measurements that can range in speed from thousands of points per second to millions. These scanners are particularly well suited to applications requiring comprehensive inspections.

As we mentioned above, these 3D scanners operate using the principle of triangulation (see Fig. 5.14). A laser point or line is first projected on the object being measured. The resulting pattern is then recorded with optical components (usually one or two cameras) based on the principle of triangulation:

The main challenge with this type of laser scanner is to accurately follow scanner movements so as to place all reconstructed points in the same referential and align or record all of the points measured.
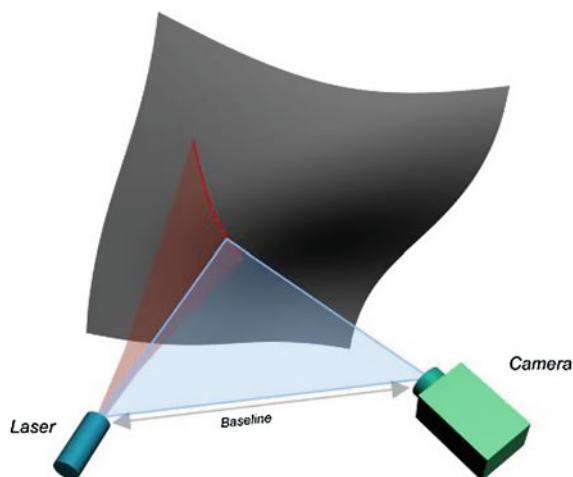
Fig. 5.14   Triangulation-based laser scanning basics

Fig. 5.15 *Left* Coordinate measuring machine (CMM). *Right* 3D laser scanner mounted on a CMM

5.4.5 Laser Scanners Using an External Positioning Device

The natural solution to this challenge was to combine laser scanners with another system that had a referencing system, in this case laser scanners using an external positioning device. With this type of system, each laser line is recorded by the laser scanner and placed in a three-dimensional coordinate system defined by the external positioning system. The practice of combining scanning systems with CMMs or portable measuring arms illustrates this concept well (see Fig. 5.15).

Other external referencing systems, such as a local GPS, may be used to obtain all the points measured in the same coordinate system. Even though these systems are all different, they nonetheless have one thing in common: To produce an accurate measurement, the referencing system they use must have a fixed position in relation to the object being measured. Consequently, special attention must be paid to the rigidity of the measuring setup when using the laser scanner with an external positioning reference (Fig. 5.16).

5.4.6 Intelligent Measurement

Intelligent measurement is the process that produces consistent measurement obtained from a set of basic observations, and which consistency is verified against a model of the measuring process (the acquisition model) during capture. Reading an ambient temperature of 60 °C in a conference room from a thermometer, or reading a diameter of 100 mm for a coin with a caliper, obviously makes no sense. One would at least make one or several additional readings before providing a reliable measurement or reporting a problem with the device. Indeed, the observation is not consistent with the expected range of values in the current context of measurement.

Fig. 5.16 Example of a rigid measuring setup when using a portable CMM

Several observations (readings) would be necessary to be able to confirm the consistency of each observation. Consistency is verified with respect to the acquisition model, which integrates knowledge on the measuring device and hypotheses on the object to be measured. This knowledge and these hypotheses may be as simple as the typical range of values and the resolution, repeatability, and accuracy of the measurement that can be produced by the device. When observations are consistent, improved measurements can also be produced after reducing error due to noise, from several observations. We will now consider that the acquisition model is integrated within the measurement process; that would yield an intelligent measuring system. In the preceding examples, the human is the intelligent agent in the measurement process.

Preferably, intelligence would be integrated into the computerized system when the quantity of observations increases and becomes intractable for the human. Integration of intelligence then becomes particularly valuable with the enormous amount of observations that can be acquired in real time with 3D optical sensors.

5.4.7 Self-positioning 3D Laser Scanners and Their Benefits

Laser scanning systems that use retroreflective targets as references need no external positioning systems, because the scanner relative position in relation to the part or object being measured is determined by the principle of triangulation
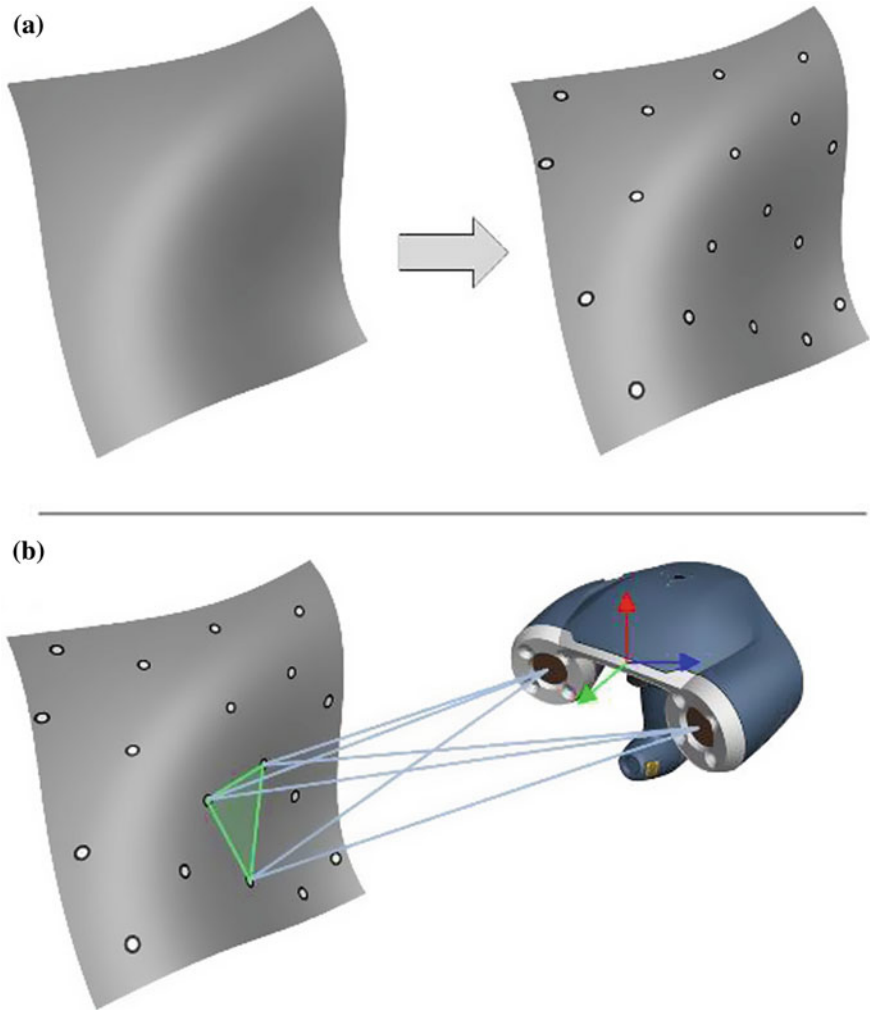
Fig. 5.17 a When a free-form surface is measured using a self-positioning system, positioning targets are affixed to the part. **b** The scanner determines its relative position in relation to the object using the principle of triangulation between the scanner two cameras and the random patterns of positioning targets

between the scanner's two cameras and the patterns of positioning targets, which is carried out in real time during data acquisition (see Fig. 5.17b).

There are three main benefits to use these scanning systems in metrology. The first applies to all scanning systems: the ability to measure complex or free-form surfaces. Three-dimensional measurement by probing is the ideal way to measure geometric shapes, but quickly proves ineffective and ill suited to measuring free-form surfaces that require a measuring speed of thousands of points per second.

Scanning systems make it possible to quickly and effectively measure a very large number of points and fully characterize a complex or free-form surface.

The second advantage concerns self-positioning 3D scanners only: They are portable, and their systems are simple to operate. Since these measuring systems do not require external references, they are much easier to use and move. As a result, they can be easily transported to the part being measured, no matter where it is.

The third advantage, and by no means the least important, is that these scanners do not require a fixed measuring setup during data acquisition—unlike other measuring systems where the system reference and part being measured must be set up on a rigid mount. For example, when measuring a part using a CMM or a measuring arm, you must make sure that:

1. The part is securely fastened to a steel or marble table or any other support capable of providing a high level of rigidity and stability;
2. The CMM or measuring arm is also fixed in relation to the part (vibrations in the measuring environment must be taken into account).

With self-positioning scanners, the measurement reference is attached directly to the object, because the positioning targets are affixed to it (see Fig. 5.17a). As a result, even if the object moves during data acquisition, the reference will remain fixed in relation to the part. These scanning systems use a reference that is measured in relation to the object on a continuous basis by the scanner's cameras during data acquisition.

5.4.8   Creating 3D Measurements

5.4.8.1   Laser Scanner Traditional Output

Once the scanner is positioned in space in relation to the part being measured, it must generate 3D measurements. At this stage, all laser scanners operate the same way by projecting one or more laser lines that conform to the shape of the object. The laser lines distorted by the object surface are then read and recorded by the scanning system camera(s) (see Fig. 5.18), and the points extracted from this image are then placed in 3D space based on the reference or coordinate system created beforehand. Once data acquisition is complete, traditional laser scanners produce a 3D point cloud (see Fig. 5.18), whereas the latest generation of portable scanners (which use automatic referencing via positioning targets) operates according to a similar process, but produces more advanced and reliable measurements.

5.4.8.2   Optimizing 3D Measurements

Self-positioning scanners also use a laser projection to characterize the surface of an object. The laser is projected in the shape of a crosshair (see Fig. 5.19a). Then, the
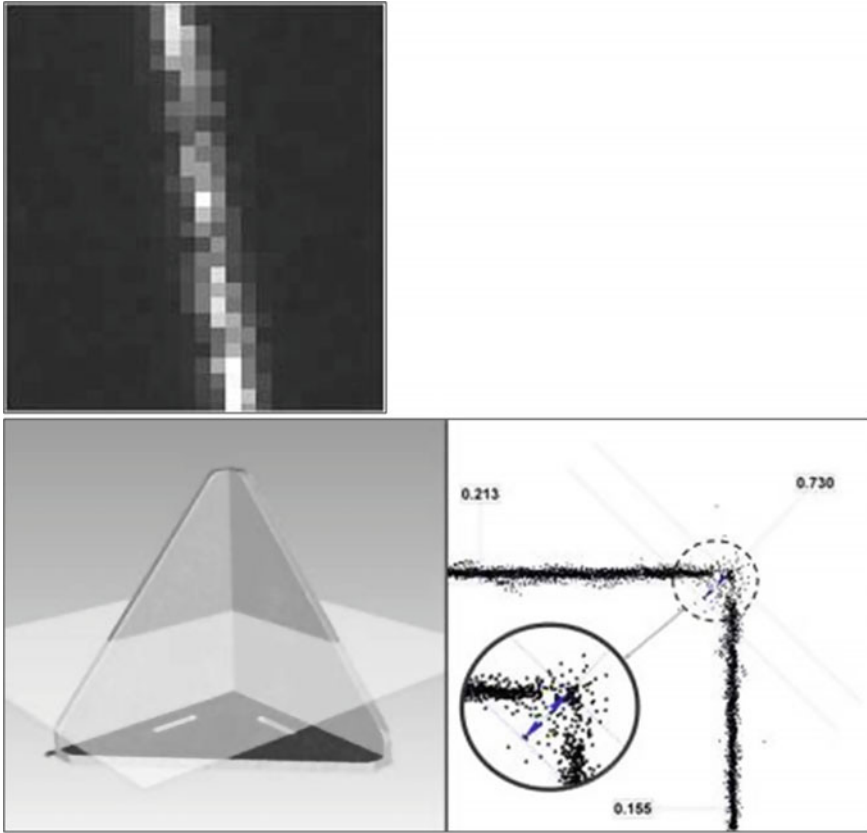
Fig. 5.18 *Top* Example of an image section of a laser trace. *Left* Example of a 3D point cloud: output of a traditional laser scanner. *Right* Distribution of the 3D points along the cross section

surface is scanned by manually moving the scanner over the entire surface to be measured (see Fig. 5.19b). However, with this type of scanner, the measurement produced is not a 3D point cloud, but rather a more highly advanced and optimized 3D measurement shape: a mesh from an optimized surface that takes into account measuring conditions and measurement consistency (see Fig. 5.19b).

5.5 Case Studies: Optical CMMs for the Aeronautical Industry

We will now illustrate how optical CMMs operate in the real world, by introducing two cases studies from the aeronautical realm. Each case study clearly identifies the benefits of using this type of measurement solution.
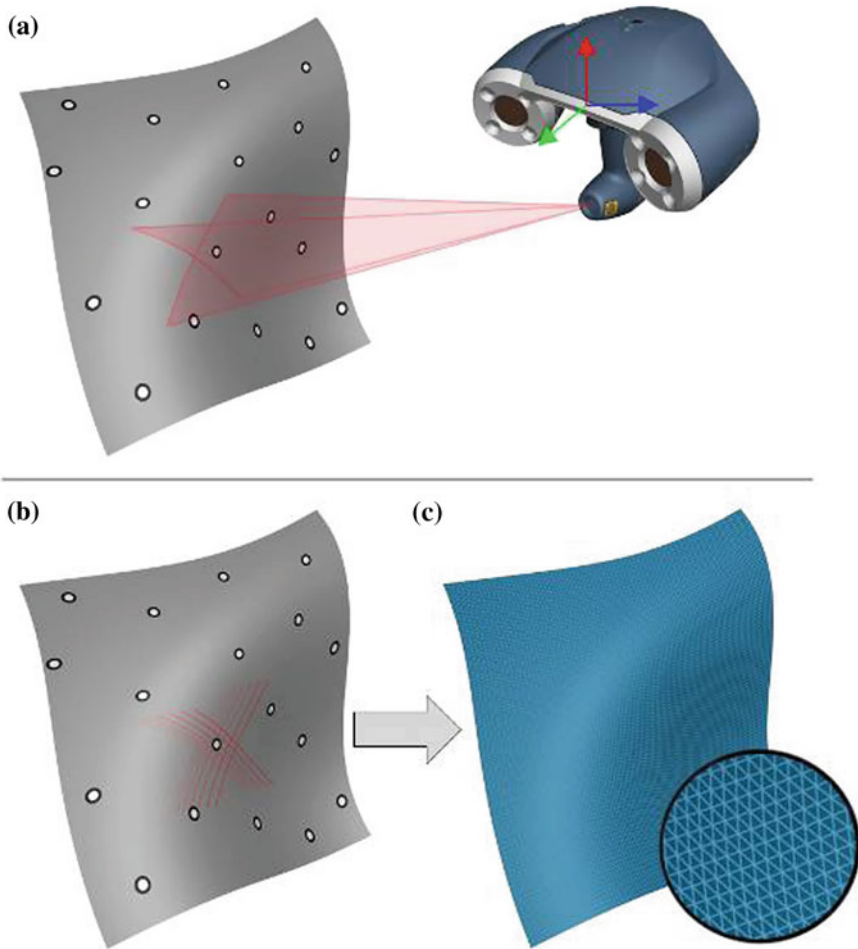
Fig. 5.19 a The self-positioning scanner projects a laser crosshair onto the surface of the object.
b The object is scanned by manually moving the scanner over the entire surface to be measured.
c An optimized mesh is generated

5.5.1 Aircelle

Aeronautical manufacturer Aircelle needed a solution to check the inverter doors of
an airplane engine. A total of 150 points had to be measured manually in less than
15 min, including the inverter door installation on the jig. The jig had to accom-
modate four types of doors, and the inspection report had to be generated
automatically.

The first step was to align the rotation axis of the door and the position of the lock. As points need to be taken on both sides of the door, the alignment had to be saved when rotating the door, for future use.

Two reflectors were fixed on the rotation axis of the door. They were automatically measured by an optical tracker, generating two points used to compute a line.

Two other reflectors were placed on the plane support for the lock. The position of the reflectors was defined to ensure that the median point between the reflectors is at same height as the center point of the lock.

Three points are now available to create an automatic alignment. Aligning the part only takes a few seconds. To make it work with different door types and in both door positions, reflectors were positioned in the different places reached by the lock. Several reflectors are also added on the jig to continuously check jig conformity.

The control program primarily includes step and gap points all around the part and is repeated on both sides of the door. Using the metrology software probing assistant, the operator simply needs to probe points at the location indicated by a target in the CAD view. Only points taken inside a sphere centered on the target point are accepted.

When the door is rotated, the positions of the reflectors placed on the jig around the lock are automatically used for the alignment. They are measured at the beginning of the process so as to become invisible when rotating the door. The measurement process can thus continue.

Some specific tools have also been developed to obtain fast and accurate measurement of specific entities. For example, a tool with a centered target was developed to measure the hole used for the hydraulic jack. The tool is positioned by the operator at the appropriate spot before starting the measurement process. By measuring the target, the center of the hole is automatically computed, eliminating the need to probe multiple points (on the supporting plane and inside the hole) and the potential errors associated with this process. Probing assistants are always available during the program to assist the operator by showing where and what to probe.

Thanks to automatic alignment, control time was drastically reduced from 2 h with a traditional CMM to only 15 min including the time required to put the door on the jig and to remove it. Control time was divided by a factor of ten.

By using automatic alignment and dynamic referencing, combined with probing assistants, 75 % of errors are avoided. An automatic continuous verification of the jig is also available thanks to the target permanently placed on the jig. The position of these targets can be verified on an annual basis using laser trackers and compatible tooling (available from major optical tooling manufacturers).

A report is automatically issued in the format specified by the customer. The operator cannot change the results or the inspection program because of an access rights system. The light wireless probe and the user-friendly system make the inspection process very comfortable for the operator (Fig. 5.20).

Fig. 5.20 Jig equipped with optical reflectors for pivot door measurement and optical portable CMM

5.5.2 Swiss Jet Aviation AG

Retroreflective targets used for dynamic referencing are also compatible with photogrammetry solutions. They make it possible to develop a comprehensive and accurate positioning model using targets affixed to a large object. Once the model is complete, an optical CMM can be placed anywhere a sufficient number of previously measured targets are visible. In this way, the optical CMM is automatically linked to the global reference. This is particularly useful for measuring very large objects (Fig. 5.21).

For a project with Swiss Jet Aviation AG, one of the world's largest business aviation services companies, the Creaform 3D Engineering Department measured and digitized the entire interior of a Boeing 737–800.

The goal was to provide Jet Aviation with a detailed representation of the aircraft's as-built fuselage, in order to make custom interior design and outfitting tasks easier. Having access to a 3D model would enable company engineers to identify equipment interference risks ahead of time and ensure equipment compatibility with the aircraft's structures and systems right from the start, thereby avoiding long and costly iterations to adjust equipment. The Creaform team of application engineers digitized the entire interior, using a combination of handheld Handyscan 3D scanners, a long-range Leica scanner, several HandyPROBE optical CMMs, and MetraSCAN optical CMM 3D scanners, as well as a photogrammetry system (Fig. 5.22).

Fig. 5.21 Simultaneous use of several pieces of equipment in a very instable environment, thanks to a dynamic positioning model built through photogrammetry
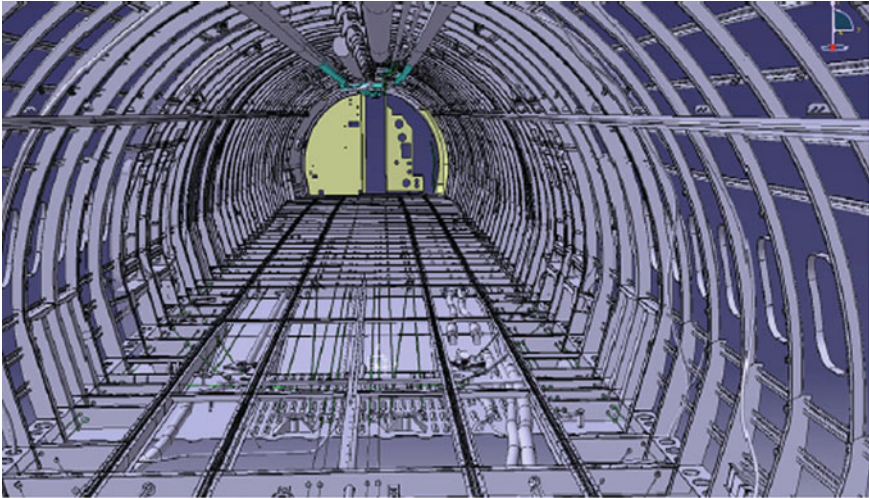
Fig. 5.22 Reconstructed CAD 3D model of the fuselage

Once the positioning model had been measured using photogrammetry, it was possible to operate all the equipment simultaneously without realigning the data afterward, and especially without worrying about vibrations generated by all the activity inside the aircraft. Once data acquisition was complete, the data were processed at Creaform and the model finalized with the CATIA software.

5.6   Conclusion

Portable optical CMMs and handheld scanners are opening a new chapter in portable 3D measurement. Their capabilities make measurement more reliable, less operator-dependent, and more portable than ever. One important benefit of this is to bring measurement directly to the shop environment production line, thereby ensuring greater reactivity during production increases, earlier detection of potential drift, and faster identification of underlying causes.

The Creaform TRUaccuracy technology—which includes dynamic referencing, automatic drift detection (through permanent control of positioning model geometry), and automatic alignment—enables users make the most of their measurement equipment and generate significant productivity gains without requiring extremely costly equipment or, most importantly, measurement environments.

Automatic alignment—and more broadly the automatic point measurement capabilities that optical CMMs already provide—is paving the way for the automation of optical measurement. Optical CMMs can already measure object deformation and movement in real time (dynamic tracking). Even though targets are still required, the scope for advancement in optics is enormous, as the rapid

development of optical scanners has shown. Clearly, the future of 3D optical measurement has only just begun.

What can be expected in the future? First, intelligent measurementis still in its infancy. Creaform has been progressively integrating this paradigm in its systems for the last six years. Currently, it is already possible to produce 3D measurements that can be exploited directly in the existing inspection software that are available commercially. In the near future, acquisition models will become more and more sophisticated and the level of intelligence integrated in the sensing systems will progressively increase as well. The acquisition systems will progress in their flexibility while remaining affordable. In the midterm, we can also expect that inspection software will be closely integrated with the sensor. These more intelligent sensors could directly accomplish inspection tasks.

References

1. CMM metrology [Web] On http://www.cmmmetrology.co.uk/history_of_the_cmm.htm. Visited on 28 Dec 2011
2. Renishaw [Web] on http://www.renishaw.com/en/Our+company–6432. Visited on 28 Dec 2011
3. AFNOR [Web] on http://www.bivi.metrologie.afnor.org/ofm/metrologie/ii/ii-80/ii-80-30. Visited on 28 Dec 2011 (in French)
4. Mayer Roy (1999) Scientific Canadian: invention and innovation from Canada's National Research Council. Raincoast Books, Vancouver. ISBN 978-1-5519-2266-9
5. Karara HM (1989) Non-topographic photogrammetry. American Society for Photogrammetry and Remote Sensing, USA
6. Atkinson KB (1996) Close range photogrammetry and machine vision. Whittles Publishing, Scotland. ISBN 1-870325-46-X
7. Hartley RI, Zisserman A (2004) Multiple view geometry in computer vision. Cambridge University Press, Cambridge
8. Bartoli A (2003) Reconstruction et alignement en vision 3D: points, droites, plans et caméras. Thèse, GRAVIR, INPG
9. Nistér D (2001) Automatic dense reconstruction from uncalibrated video sequences. Rapport de Thèse, KTH, Université de Stockholm
10. Lhuillier M, Quan L (2005) A quasi-dense approach to surface reconstruction from uncalibrated images. IEEE Trans Pattern Anal Mach Intell 27(3):418–433
11. Royer É et al (2007) Monocular vision for mobile robot localization and autonomous navigation. Int J Comput Vision 74(3):237–260 (Springer Science)
12. Agarwal S et al (2010) Bundle adjustment in the large. In: European conference on computer vision, Springer, Berlin, pp 29–42
13. Triggs B et al (2000) Bundle adjustment—a modern synthesis. Triggs B, Zisserman A, Szeliski R (eds) Vision algorithms: theory & practice. Springer, Berlin, pp 153–177
14. Jian, Yong-Dian, Balcan, Doru C. et Dellaert, Frank. 2011. Generalized subgraph preconditioners for large-scale bundle adjustment. In: *Proceedings of 13th IEEE international conference on computer vision.* IEEE, NJ
15. Morris D (2001) *Gauge freedoms and uncertainty modeling for 3d computer vision.* Thèse de doctorat, Robotics institute, Carnegie Mellon University
16. Faugeras O (1993) Three-dimensional computer vision. MIT Press, Cambridge

17. Shortis MR, Clarke TA (1994) A comparison of some techniques for the subpixel location of discrete target images, *Vdeometrics III* 2350:239–250
18. Shortis MR, Clarke TA, Robson S (1995) Practical testing of the precision and accuracy of target image centring algorithms. *Videometrics IV* 2598:65–76
19. Otepka J (2004) Precision target mensuration in vision metrology. Technische Universität Wien, Vienne
20. Hartley RI, Sturm P (1997) Triangulation. Comput Vision Image Understand 68(12):146–157
21. Horn BKP (1987) Closed-form solution of absolute orientation using unit quaternions. J Opt Soc America 4:629–642
22. Eggert DW, Lorusso A, Fisher RB (1997) Estimating 3-D rigid body transformations: a comparison of four major algorithms Machine Vision and Applications 9(5–6):272–290
23. Wikipedia [Web] on http://en.wikipedia.org/wiki/Structured-light_3D_scanner. Visited on 28 Dec 2011
24. Fringe (2005) The 5th international workshop on automatic processing of fringe patterns. Springer, Berlin. ISBN 3-540-26037-4; ISBN 978-3-540-26037-0
25. Wikipedia [Web] on http://en.wikipedia.org/wiki/File:Laserprofilometer_EN.svg. Visited on 28 Dec 2011

Chapter 6
Fast Three-Dimensional Shape Inspection Using a Multi-sided Mirror

Idaku Ishii

Abstract This chapter describes a fast and high-space resolution three-dimensional (3-D) shape inspection system based on the light-section method, which sets a multi-sided mirror between an image sensor and the target object being measured. Assuming that planar objects are being measured, the 3-D shape measurement system can enhance the width of the measurement area in a sideward direction by introducing a spatial transform of the measurement area using a multi-sided mirror. The mirror divides the image area of $h \times v$ pixels into n areas of $h' \times v$ pixels ($h = nh'$) and lines up these areas in a sideward direction. Thus, it is possible to achieve a wider range of $h' \times v'$ pixels ($v' = nv$) and accurate 3-D shape measurements in a sideward direction. A prototype system was developed for real-time 3-D shape measurement based on high-speed vision hardware that could process a 1024×1024 image at 1000 fps. To accelerate the 3-D profile calculations, a hardware circuit logic for the light-section method was implemented using eight-divided image areas of 128×1024 pixels on a field-programmable gate array (FPGA) device in the high-speed vision hardware, which allowed a 3-D profile with 8192 points to be outputted in real time. The performance was verified by measuring 3-D shape information for a 94 mm × 147 mm printed circuit board for a mobile phone.

Contents

I. Ishii (✉)
Faculty of Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan
e-mail: iishii@robotics.hiroshima-u.ac.jp

6.1   Introduction

Recently, there has been a high demand for real-time three-dimensional (3-D) shape measurement in many applications such as human modeling, industrial inspection of products, object recognition, and 3-D map building. The light-section method is a well-known optical method for acquiring accurate 3-D shape information without any complicated processing, and many real-time 3-D shape measurement systems have been developed based on the light-section method. However, most of them use general-purpose image sensors, which use standard video signals (e.g., NTSC 30 fps, PAL 25 fps), so their frame rate and spatial resolution restrict the data rate, measurement accuracy, and measurable range during 3-D shape measurement.

To address these limitations of 3-D shape measurement with general-purpose image sensors, this chapter introduces a novel 3-D shape measurement method that provides a wider measurement range than the spatial resolution of an image sensor based on the light-section method. A wider measurable range is achieved by spatial transforms of the image area of the image sensor using a multi-sided mirror. Based on the 3-D shape measurement method using a multi-sided mirror, a high-frame-rate (HFR) 3-D shape measurement system was developed using high-speed vision hardware, which could process a 1024×1024 image at 1000 fps in real time. The effectiveness of the newly developed HFR 3-D shape measurement system was demonstrated by measuring the 3-D shapes of thousands of submillimeter-height solders simultaneously on a printed circuit board, which was moved unidirectionally at a fixed speed of 17 mm/s.

6.2   Light-Section Method

6.2.1   Related Work

The light-section method [1, 2] is a vision-based 3-D shape measurement method, which can measure the 3-D shape of a target object using triangulation by projecting a slit light pattern onto the target object and processing an image of the

target object captured by a camera installed in a different direction compared with the projection direction of the slit light pattern. Using the light-section method, 3-D shapes can be obtained accurately with a high-directive light pattern projector such as a slit laser, which is not strongly dependent on surface texture patterns. Thus, the light-section method has been applied to practical 3-D shape measurement systems to inspect precise product parts such as electronic boards and press-processed parts, which often require a spatial accuracy in the order of 10 μm for the satisfactory refinement of parts according to recent studies [3, 4]. However, the number of measurement points is often insufficient for fast 3-D shape measurement when a general-purpose image sensor is used for the light-section method because only the 3-D intersected profile generated by a slit line can be outputted at a limited frame rate, i.e., dozens of frames per second.

Recent improvement of integration technology has used newly developed real-time HFR vision systems that operate at frame rates of 1000 fps or more, such as digital vision chips [5–7] and field-programmable gate array (FPGA)-based high-speed vision systems [8–11]. In particular, many HFR smart sensors have been designed specifically for 3-D shape measurement based on the light-section method because its 3-D profile calculations can be achieved by simple processing, which is suitable for circuit integration. Oike et al. [12, 13] developed a smart sensor that could obtain 3-D information at 40,000 fps or more by selecting and processing only the pixels required for depth calculation. Lindgren et al. [14] developed a 1536×512 image sensor where single-instruction multiple-data (SIMD) processors that operated at 100 giga operations per second (GOPS) were integrated to facilitate rapid 3-D profile calculations, which could measure 3-D shapes at a maximum rate of 40,000 fps.

6.2.2 Problems with Conventional Methods

The aforementioned 3-D smart sensors can greatly accelerate the speed of 3-D shape measurement, but their spatial resolution is still restricted in the vertical or horizontal directions; that is, the number of measurement points is limited to 2000 points per profile or less in the light-section method, even when using a large image sensor with four megapixels. Thus, these smart sensors are not always sufficient to achieve rapid 3-D shape measurements with a wide measurable range and high measurement accuracy. In recent years, it has often been necessary to conduct in-depth inspections of parts in a production line, such as electronic boards and press-processed parts, where rapid 3-D shape measurements are required to measure an area of dozens of centimeters with a measurement accuracy of dozens of μm in a defined task time. This requires a 3-D intersected profile measurement of 10,000 pixels or more in the vertical or horizontal directions, which is too wide for a single image sensor to cover.

To provide a wide measurable range and high measurement accuracy during rapid 3-D shape measurement, we might consider several methods such as

mechanical scanning of the target object in a different direction from the conveyor line or enhancing the measurable range by lining up multiple 3-D shape measurement systems. However, these ideas are not always suitable for actual inspections of production lines because time loss is caused by mechanical scanning in the first case, while the inspection system becomes complex and expensive in the latter case.

6.3 3-D Shape Measurement Using a Multi-sided Mirror

To obtain 3-D intersected profile of large numbers of pixels using a single image sensor, this chapter introduces an improved light-section method that can enhance the width of the measurable area in the vertical direction against a conveyor line in 3-D using a multi-sided mirror. The multi-sided mirror was constructed from several mirrors where the angles were different, which was installed between an image sensor and the target object being measured to transform the measurable area to allow 3-D shape measurements.

Figure 6.1 illustrates the concept of the spatial transform of a measurable area using an n-sided multi-sided mirror where the size of an image sensor is $h \times v$ pixels. The multi-sided mirror divides the image sensor's area into n areas of $h' \times v$ pixels ($h = nh'$), and these areas are lined up in the vertical direction against a conveyor line. n measurable areas distributed in the vertical direction relative to the conveyor line are collected using $h \times v$ image sensors, and the collected image is equivalent to an $h' \times v'$ image ($v' = nv$). As a result, the number of measurement points on a 3-D intersected profile is n times of that of conventional light-section methods, so rapid 3-D shape measurements can be conducted with a wider measurable range in the vertical direction relative to a conveyor line. For example,
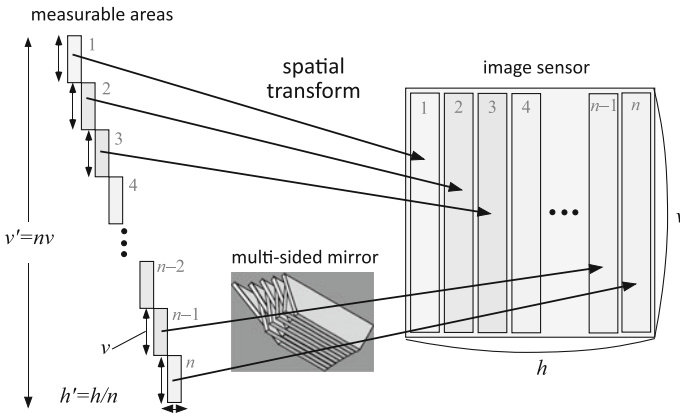
Fig. 6.1 Spatial transforms of the measurement areas using a multi-sided mirror

when an eight-sided mirror is used with an image sensor of 1024 × 1024 pixels, the image acquired is equivalent to the measurable area of an image sensor of 128 × 8192 pixels and a 3-D intersected profile of at most 8192 pixels can be measured using the image obtained for every image frame.

An advantage of this method is that it allows wide-range and high-accuracy 3-D shape measurements when the 3-D shape of the target object being measured is inspected on a conveyor line. There is a trade-off between the measurable range in the vertical direction relative to the conveyor line and the measurement accuracy in the depth direction because the measurable range in the same direction as the conveyor line becomes $1/n$ compared with the case when a single image sensor is used for the light-section method. Thus, a 3-D shape measurement system can be installed easily for inline inspection without large-scale modifications to existing production lines simply by moving the target object in the direction of the conveyor line without any mechanical scanning.

6.4   HFR 3-D Shape Measurement System

6.4.1   System Overview

Based on the concept of the improved light-section method using a multi-sided mirror introduced in the previous section, this section describes a prototype HFR 3-D shape measurement system for planar objects on a conveyor line. Figures 6.2
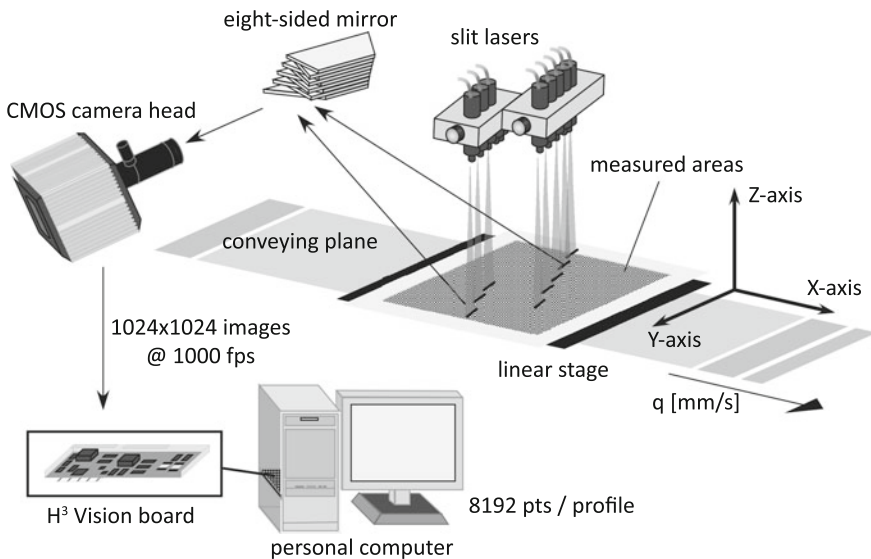
Fig. 6.2 Framework of the HFR 3-D shape measurement system
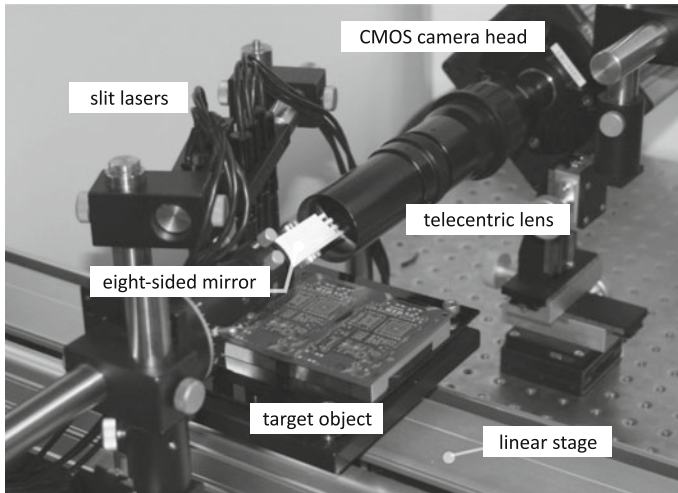
Fig. 6.3 Overview of the HFR 3-D shape measurement system

and 6.3 show the framework of the 3-D shape measurement system and its overview, respectively. This system comprises optics, i.e., the multi-sided mirror, slit lasers, and a telecentric lens, and an H^3 Vision system [10] that can process 1024×1024 slit images in real time at 1000 fps. The slit images are transformed spatially using the multi-sided mirror. The target object being measured is conveyed in a single direction at $q = 17$ mm/s using an electric linear stage (NSK, XY-HRSO50-PH200).

6.4.2 Optics Including a Multi-sided Mirror

Figure 6.4 shows the composition of the optics in this 3-D shape measurement system. With respect to the *XYZ* coordinates used during 3-D shape measurements, the *X*-axis is the direction in which a target object is conveyed and the *Y*-axis is at right angles to the conveying plane, while the *Z*-axis is vertical to the conveying plane. Figure 6.4a shows a side view from the *X*-axis direction, and Fig. 6.4b shows the top view from the *Z*-axis direction.

A telecentric lens (Carl Zeiss, 16/16/0.1-LD) with a magnification ratio *T* of 1 and a working distance of 92 mm is mounted on the camera head of the H^3 vision system. The camera head uses a 1024×1024 CMOS image sensor with a size of 17 mm $\times$ 17 mm, and the image range captured is also 17 mm $\times$ 17 mm on the conveying plane in the working distance.

The image range of the image sensor is divided by the eight-sided mirror shown in Fig. 6.5a. Figure 6.5b and c shows the dimensions of the eight-sided mirror. Figure 6.5d shows the dimensions of the individual mirrors that comprise the
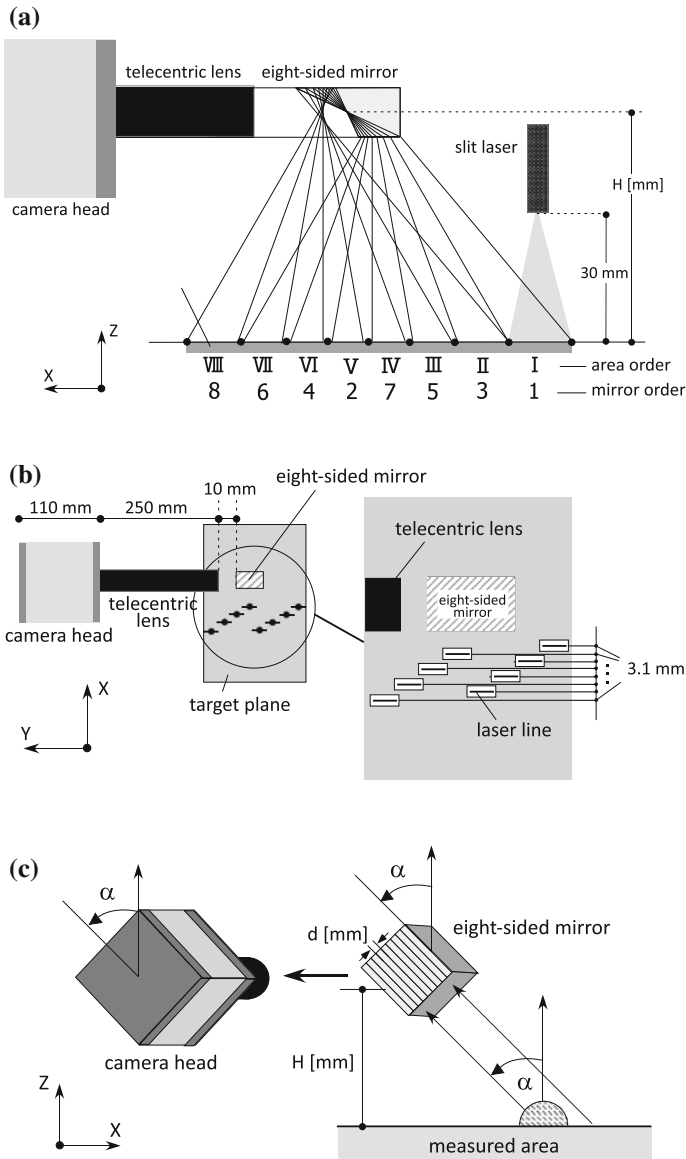
Fig. 6.4 Configuration of the optical system: **a** side view, **b** top view, and **c** orientation

eight-sided mirror. The eight-sided mirror was constructed by gluing eight trape-zoidal aluminum mirrors with different angles. For all of the mirrors, the height is 20 mm and the thickness $d = 2.176$ mm where the thickness of each mirror corresponds to 128 pixels of the image sensor, assuming that they divide the view of the telecentric lens into eight parts. Each mirror is placed at a location that
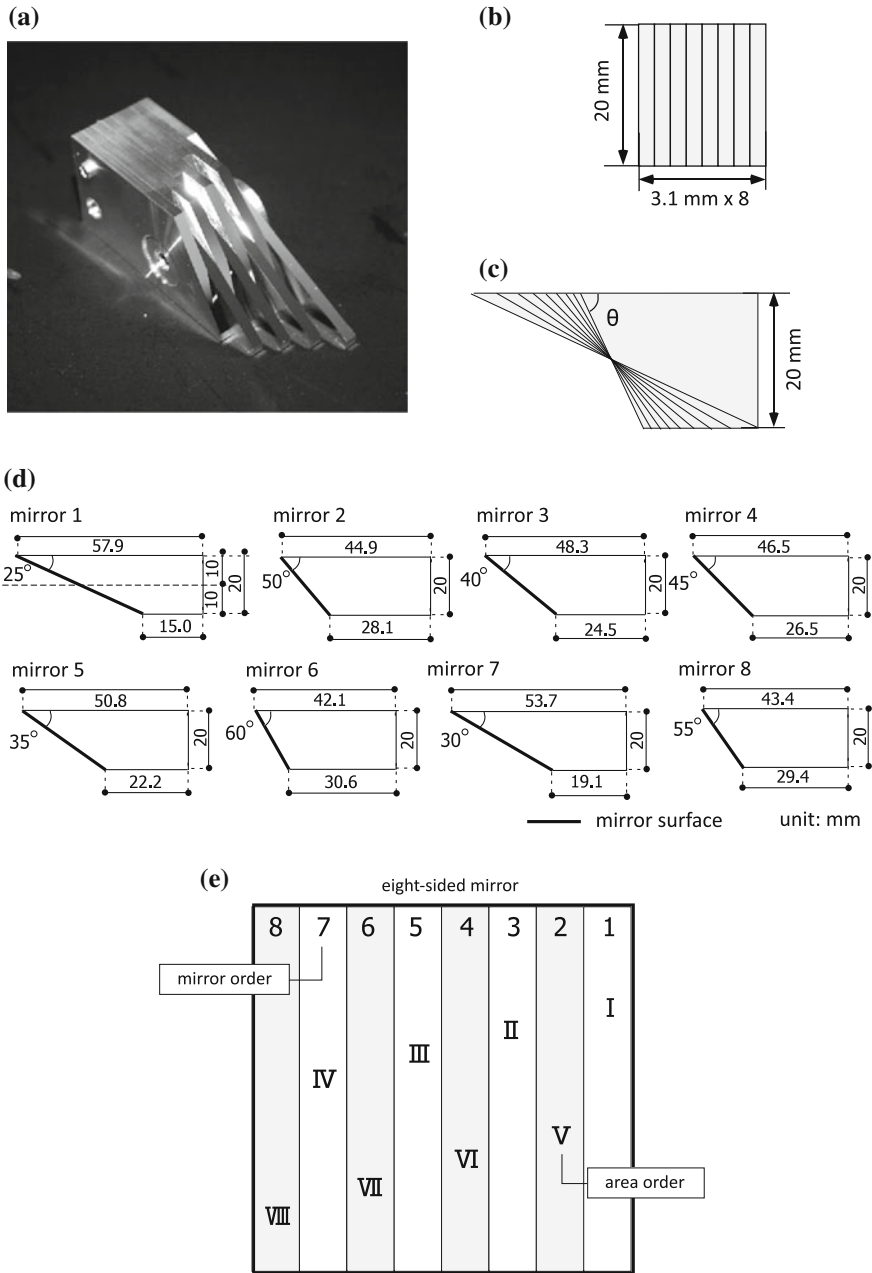
Fig. 6.5 Configuration of the eight-sided mirror: **a** overview, **b** front view, **c** side view, **d** scales, and **e** image areas allocated to the mirrors

corresponds to a rectangular area of 128×1024 pixels in the image sensor. All of the mirror surfaces are glued so they share identical straight lines at a height of 10 mm, as shown in Fig. 6.5c.

The angle of each mirror surface was designed so the width of the measurable area was maximized at a distance of 95 mm from the central position of the mirror. The angles θ_m for mirrors 1, 2, …, and 8 are set to 25°, 45°, 30°, 50°, 35°, 55°, 40°, and 60°, respectively. Figure 6.4a shows the measurement areas that correspond to mirror numbers 1, 2, …, and 8 lined up in the Y direction in the following order: 1, 3, 5, 7, 2, 4, 6, and 8. The order of the mirror angles was determined so the measurement areas of adjacent mirrors did not overlap because of reflections of the slit lasers when viewed from above. Thus, the measurement areas are projected onto the rectangular areas above the image sensor of the camera head in the order shown in Fig. 6.5e.

The eight-sided mirror and the camera head are set in a position so the center of the mirror is at a height of $H = 40$ mm from the target plane and tilted by approximately $\alpha = 45°$ in the Z direction, which is the same as that of the slit projector. Red semiconductor lasers at a wavelength of 635 nm (Kikoh Giken, MLXH-A13-655-50) are used as the slit illuminators for these mirrors, which can project a smallest slit line of 13 μm. These slit lasers are arranged at every $d/\cos\alpha = 3.1$ mm in the X direction, which corresponds to the thickness of the mirror, as shown in Fig. 6.4b. Each slit laser projects slit lines from 30 mm above and perpendicular to the target plane.

6.4.3 High-Speed Vision Platform, H^3 Vision

The H^3 vision system [10] is used as a high-speed image processing platform for gray-level 1024×1024 images captured at 1000 fps. Figure 6.6a shows the configuration of this platform. This platform comprises a high-speed camera head for imaging a target object, a dedicated FPGA image processing board (H^3 vision board) that can be coded for image processing functions in hardware logic, and a personal computer (PC).

The FASTCAM-1024PCI (Photron) is used as a high-speed camera head, which can capture and transfer 10-bit gray-level images of 1024×1024 pixels at 1000 fps to a post-stage H^3 Vision board. Figure 6.6b shows an overview of the system. There is a CMOS image sensor that measures 17 mm × 17 mm with a pixel pitch of $a = 0.017$ mm. The H^3 Vision board is designed to be a dedicated FPGA board for high-speed image processing. Figure 6.6c and Table 6.1 show its overview and specifications, respectively. This board comprises the FPGA (Xilinx XC2VP100, referred to as FPGA1) used for image correction and for displaying the processed images, the FPGA (Xilinx XC2VP100, referred to as FPGA2) for hardware logic implementation of the image processing algorithms by the user, DDR-SDRAM and various other memories, and interface circuits such as a PCI-X bus and bridge. Using the H^3 Vision board, a 10-bit gray-level 1024×1024 image can be processed
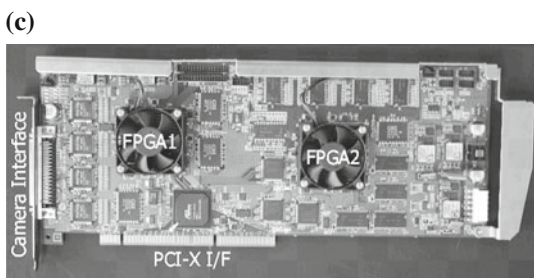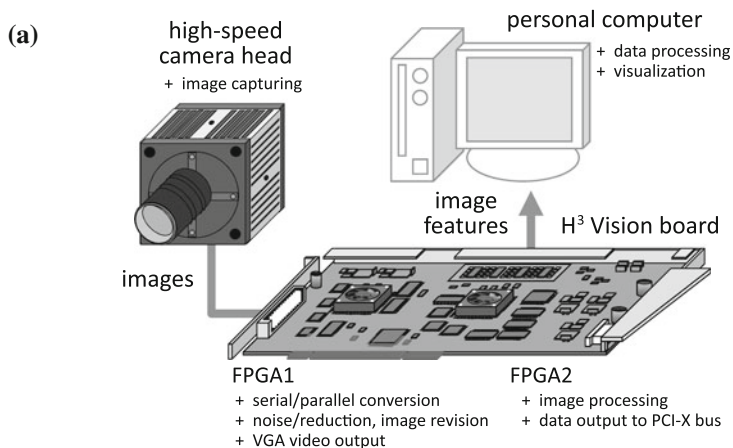
Fig. 6.6 High-speed vision platform, H³ Vision: **a** system configuration, **b** camera head, and **c** H³ Vision board

Table 6.1 Specifications of the H³ Vision board

Camera-I/F	Digital serial 12ch Color/gray 10 bit
Image adjustment	Fixed pattern noise reduction Shading compensation
Memory (FPGA2)	DDR-SDRAM (640 MB) SDRAM (512 MB) SSRAM (9 MB) FPGA internal memory (999 KB)
FPGA	XC2VP100 (Xilinx) × 2
PCI-I/F	PCI-X (64 bit, 66 MHz) FPGA internal local bus (32 bit, 66 MHz)

and stored in real time at 1000 fps using a 160-bit bus between FPGA2 and DDR-SDRAM. Figure 6.7 shows a function block diagram of the H³ Vision board. The image data path in the H³ Vision board is described as follows:
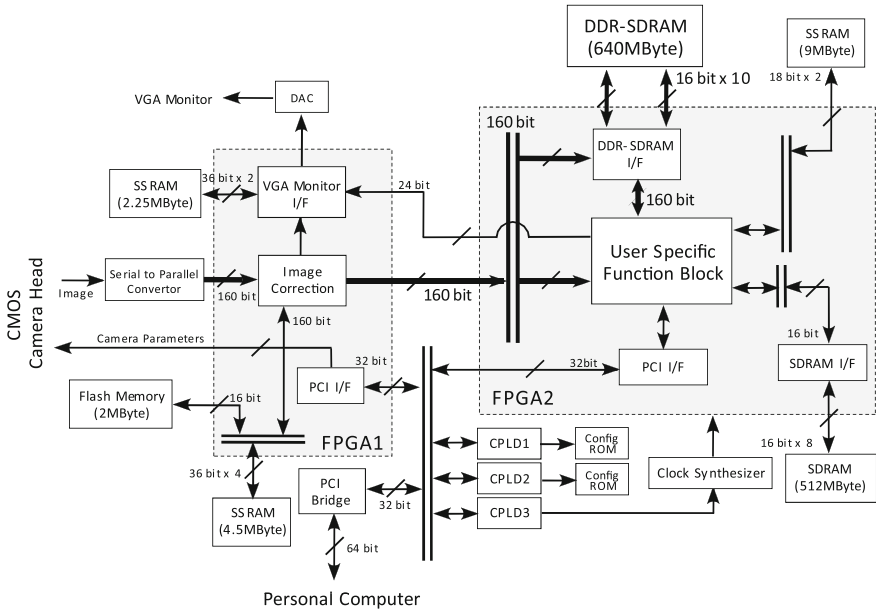
Fig. 6.7 Block chart of the H^3 Vision board

1. Converting image data from serial to parallel
   Serial image data transferred from the high-speed camera head are converted into 160-bit parallel data (10 bits/pixel) with 16 pixels as a parallel block.
2. Correcting images
   Using prestored corrected images in the flash memory, fixed pattern noise reduction and shading correction are conducted for the parallel image data from (1).
3. Writing corrected images to DDR-SDRAM
   The corrected image data are transferred once from FPGA1 to FPGA2 via a 160-bit bus and stored in DDR-SDRAM, where the DDR-SDRAM functions as a buffer.
4. Executing user-specified image processing
   Stored images are read from DDR-SDRAM and processed in a processing block where arbitrary user-implemented image processing is executed in FPGA2.
5. Transferring the processing results to the PC
   Using high-speed transfer via FIFO, the processing results from (4) are sent to the PCI-X bus through the PCI-I/F and PCI bridge for data acquisition by the PC.

To facilitate rapid 3-D shape measurement, a hardware logic function that calculates the centroids of the row direction for 3-D shape measurement based on the light-section method for each divided image area of 128 × 1024 pixels, where a 1024 × 1024 image is divided by the eight-sided mirror, is implemented on the

Table 6.2 Resource consumption of FPGAs

	FPGA1	FPGA2
Slice	4048/44,096 (9 %)	36,929/44,096 (83 %)
Input/output	955/1164 (82 %)	1024/1164 (87 %)
Block RAM	71/444 (15 %)	345/444 (77 %)

FPGA2 on the H^3 Vision board. The centroids of the row directions for each divided image area, $^mC_x(y,t)$ $(m = 1, 2, \ldots, 8, y = 1, 2, \ldots, 1024)$, are obtained after calculating the zeroth-order moment and the first-order moment of the row direction, $^mM_0(y,t)$ and $^mM_1(y,t)$, for each divided image area as follows:

$$^mM_0(y,t) = \sum_{x=128(m-1)+1}^{128m} I(x,y,t), \tag{6.1}$$

$$^mM_1(y,t) = \sum_{x=128(m-1)+1}^{128m} xI(x,y,t), \tag{6.2}$$

$$^mC_x(y,t) = \frac{^mM_1(y,t)}{^mM_0(y,t)}. \tag{6.3}$$

where $I(x,y,t)$ is the binary image at time t. The centroid value calculated corresponds to the height of the Z direction, which is based on the light-section method. As a result, the 8192-point 3-D profile can be determined from a frame image because 1024 points of the centroid are calculated for eight-divided images of 128×1024 pixels. Table 6.2 shows the resource consumption of FPGAs where the hardware logic can calculate the centroids of the row directions for eight 128×1024 divided image areas, which is implemented on the H^3 Vision board. We confirmed that this process could be performed in real time at 1000 fps and the centroid values calculated were transferred to the PC as 8-bit integers. The PC used in this implementation was based on the Intel Server Board SE7525GP2, which has a PCI-X bus, a Xeon 2.8 GHz CPU, 2 GHz memory, and the OS was Windows XP.

6.4.4 Calculation of 3-D Shape Information

Using the PC, the 3-D position that corresponds to the pixel position y at time t in each image area $m(= 1, 2, \ldots, 8)$ divided by the eight-sided mirror can be calculated using the image centroid $^mC_x(y,t)$ of the row direction as follows:

$$^m\mathbf{P}(y,t) = (^mP_X(y,t), ^mP_Y(y,t), ^mP_Z(y,t))^T. \tag{6.4}$$

Figure 6.8 shows the geometrical relationship between $^m\mathbf{P}(y,t)$ and the XZ-plane, the XY-plane, and the YZ-plane of the 3-D shape measurement system.

Based on the position of the slit laser projected onto the m_0th divided image area at time $t = 0$, the X coordinate $^mP_X(y,t)$ at time t and the pixel position y when the target object being measured is conveyed at a uniform speed q can be obtained based on the interval of the slit lasers, $d/\cos\alpha$, as follows:

Fig. 6.8 Measurement parameters: **a** XZ-plane, **b** XY-plane, and **c** YZ-plane
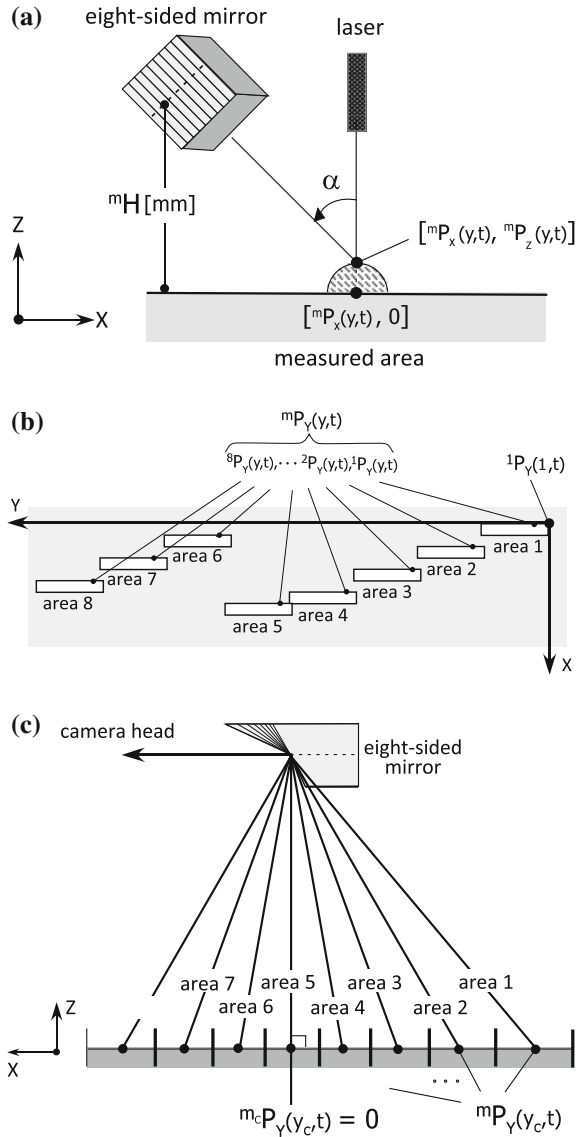
$$^{m}P_X(y, t) = qt + \frac{(m - m_0)d}{\cos \alpha}. \tag{6.5}$$

The Y coordinate $^{m}P_Y(y, t)$ at time t and the pixel position y can be obtained using the following procedure.

Figure 6.8c shows that the slit laser projected at the center of the image $y = y_c$ in all the divided image areas passes along the identical straight line shared by all of the mirror surfaces. Depending on the inclination α of the mirror, the height ^{m}H of the center of the mirror surface that corresponds to the mth divided image area is different from that of the other mirror surfaces, as shown in Fig. 6.8a, which can be expressed as follows:

$$^{m}H = (m_1 - m)d \sin \alpha + H, \tag{6.6}$$

where H is set to the height of the center of the mirror surface, which corresponds to the m_1th divided image area. The Y coordinate $^{m}P_Y(y_c, t)$ that corresponds to the coordinate of the center position $y = y_c$ in the m_cth divided image area, where the light reflected by the mirror and the measurement plane cross vertically, is assumed to be zero. Thus, $^{m}P_Y(y_c, t)$ can be computed using the angle θ_m of the mirror surface, which corresponds to the m-th divided image area, as follows:

$$^{m}P_Y(y_c, t) = -\frac{^{m}H}{\sin \alpha \cdot \tan 2\theta_m}. \tag{6.7}$$

Thus, $^{m}P_Y(y, t)$ at time t and pixel position y, including $^{m}P_Y(y_c, t)$ as a baseline, can be derived approximately by using the magnification ratio T of the telecentric lens and the pixel pitch a of the image sensor as follows:

$$^{m}P_Y(y, t) = {}^{m}P_Y(y_c, t) - \frac{a(y - y_c)}{T \sin 2\theta_m}. \tag{6.8}$$

Based on the principle of the light-section method, the Z coordinate $^{m}P_Z(y, t)$ at time t and the pixel position y can be calculated using the centroid $^{m}C_x(y, t)$ of the row direction as follows:

$$^{m}P_Z(y, t) = \frac{a}{T \cos \alpha} {}^{m}C_x(y, t). \tag{6.9}$$

Assuming that the slit lasers project extremely thin lines onto the target object being measured, the error $\Delta^{m}\mathbf{P}(y, t)$ of 3-D shape measurement expressed using Eqs. (6.5), (6.8), and (6.9) is determined as follows:

$$\Delta^{m}\mathbf{P}(y, t) = \left(\frac{q}{F}, \frac{a}{T \sin 2\theta_m}, \frac{a}{T \cos \alpha} \right)^{T}, \tag{6.10}$$

where F is the frame rate of the image sensor and the digitized error when calculating $^mC_x(y_t)$ is assumed to be one pixel.

The measurable range in the X direction depends on the conveying area of the target object being measured. When an $h \times v$ image sensor is divided into n image areas by the multi-sided mirror, each measurable range in the Y direction and the Z direction can be expressed as follows:

$$^1P_Y(1,t) \leq {}^mP_Y(y,t) \leq {}^nP_Y(v,t) \tag{6.11}$$

$$0 \leq {}^mP_Z(y,t) \leq \frac{a}{T\cos\alpha}\left(\frac{h}{n}-1\right), \tag{6.12}$$

where h/n is the number of horizontal pixels in the divided image area. We assume that the measurable ranges of the neighbor-divided image areas overlap in the Y direction around their boundaries.

In the prototype system, the 3-D position $^m\mathbf{P}(y,t)$ expressed with Eqs. (6.5), (6.8), and (6.9) is calculated in real time using the following parameters:

$q = 17$ mm/s, $d = 2.176$ mm, $\alpha = 45°$
$H = 40$ mm, $a = 0.017$ mm/pixel, $T = 1$
$F = 1000$ fps, $n = 8$, $h = v = 1024$ pixels
$m_0 = 1$, $m_1 = 8$, $m_c = 2$, $y_c = 512$ pixels
$(\theta_1, \ldots, \theta_8) = (25°, 45°, 30°, 50°, 35°, 55°, 40°, 60°)$

With these settings, the 3-D position $^m\mathbf{P}(y,t)$ is expressed in millimeters as follows:

$$^mP_X(y,t) = 0.017t + 3.07(m-1), \tag{6.13}$$

$$^mP_Y(y,t) = {}^mP_Y(512,t) + \frac{0.017}{\sin 2\theta_m}(y-1), \tag{6.14}$$

$$^mP_Z(y,t) = 0.024\,{}^mC_x(y,t). \tag{6.15}$$

Thus, the measurement error is 0.017 mm in the X direction, 0.017–0.022 mm in the Y direction, and 0.024 mm in the Z direction. The measurable range of the m-th divided image area is expressed in millimeters as follows:

$$0.000 \leq {}^1P_Y(y,t) \leq 22.724, \tag{6.16}$$

$$16.118 \leq {}^2P_Y(y,t) \leq 36.219, \tag{6.17}$$

$$28.977 \leq {}^3P_Y(y,t) \leq 47.502, \tag{6.18}$$

$$40.016 \leq {}^{4}P_Y(y,t) \leq 57.692, \tag{6.19}$$

$$50.124 \leq {}^{5}P_Y(y,t) \leq 67.532, \tag{6.20}$$

$$59.965 \leq {}^{6}P_Y(y,t) \leq 77.641, \tag{6.21}$$

$$70.155 \leq {}^{7}P_Y(y,t) \leq 88.680, \tag{6.22}$$

$$81.438 \leq {}^{8}P_Y(y,t) \leq 101.539, \tag{6.23}$$

where the *XYZ* coordinate system is set so ${}^{1}P_Y(1,t) = 0$, as shown in Fig. 6.8a. The measurable ranges ${}^{m}P(y,t)$ in the *Y* direction overlap around their boundaries to ensure continuous measurement in the *Y* direction. Thus, the 3-D shape measurement system has a measurable range of 0.000–101.539 mm in the *Y* direction and 0.000–3.053 mm in the *Z* direction. The *Z* coordinate is calculated as the average value of the *Z* coordinates measured in the neighbor image areas when they had overlapped areas in the *Y* direction.

6.4.5   Data Correction

The equations described in Sect. 6.4.4 assume that all of the slit laser patterns projected in the divided image areas correspond to ${}^{m}C_x(y,t) = 0$, which is the baseline for 3-D shape measurement, when the slit laser patterns are projected onto the conveying plane. In a real setting, however, the divided image areas have small geometrical errors between the target object being measured and the slit laser patterns projected onto the target object, which depend on alignment errors in the optics and mechanical errors in the conveyor control system.

The geometrical errors in the divided image areas can be reduced using a data correction process, which utilizes the slit laser patterns projected onto a flat plane such as the base plate that attaches the target object to a conveyor. During the data correction process, the centroid ${}^{m}C_x(y,t)$ of the row direction is obtained for a uniform-height base plane and the approximate straight line ${}^{m}G_x(y,t)$ is estimated by applying the least squares method to the centroid ${}^{m}C_x(y,t)$ of the row direction in the divided image area, as shown in Fig. 6.9:
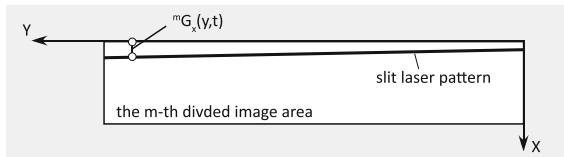
Fig. 6.9   Slit laser's position projected onto a base plane

$$^{m}G_{x}(y,t) = \frac{y}{^{m}e(t)} - {}^{m}f(t). \tag{6.24}$$

where the line parameters $^{m}e(t)$ and $^{m}f(t)$ are given as follows:

$$^{m}e(t) = \frac{1024 \sum_{y=1}^{1024} y^{m}C_{x}(y,t) - \sum_{y=1}^{1024} y \sum_{y=1}^{1024} {}^{m}C_{x}(y,t)}{y \sum_{y=1}^{1024} {}^{m}C_{x}^{2}(y,t) - \left(\sum_{y=1}^{1024} {}^{m}C_{x}(y,t)\right)^{2}}, \tag{6.25}$$

$$^{m}e(t) = \frac{1}{1024} \left(\sum_{y=1}^{1024} y - e \sum_{y=1}^{1024} {}^{m}C_{x}(y,t)\right). \tag{6.26}$$

The corrected Z coordinate $^{m}P_{Z}(y,t)$ can be obtained by substituting the difference between the approximate straight line $^{m}G_{x}(y,t)$, which corresponds to the base plane, and the centroid $^{m}G_{x}(y,t)$ of the row direction, when observing the target object into Eq. (6.9), as follows:

$$^{m}P'_{Z}(y,t) = \frac{a}{T\cos\alpha} \{^{m}C_{x}(y,t) - {}^{m}G_{x}(y,t)\}. \tag{6.27}$$

6.5 Experimental Results

6.5.1 Accuracy Verification

First, the measurement accuracy in the Z direction was verified using the 3-D shape measurement system described in Sect. 6.4. In the experiment, a white acrylic plate was used as the target object being measured. After mounting the plate on a z-movable stage, its height was changed over a range of 200 μm at an interval of 10 μm based on a height of $H = 40$ mm. The alignment accuracy of the z-movable stage was 5 μm, and the white acrylic plate had no displacement in the X direction during the experiment. It was confirmed that there were no light reflections from inside the white acrylic plate when the slit lasers were projected onto it.

As an example, Fig. 6.10 shows the relationship between the actual heights of the white acrylic plate and the heights measured on a line of $y = 512$ in the 4th divided image area ($m = 4$). It is considered that the measurement error was caused mainly by the digitized noises of the image sensors when measuring the positions of the slit laser patterns projected on the target object. In the experiment, the width of the slit laser projected on the target object corresponded to approximately one pixel in the divided image region, so the accuracy when calculating the centroid of the row direction was at the integer level, i.e., the measurement error depended on the change in height per pixel. Thus, discrete errors can be seen every 24 μm in Fig. 6.10, which correspond to Eq. (6.15). Such measurement errors in the

Fig. 6.10 Relationship
between the actual height and
the measured height ($m = 4$,
$y = 512$)
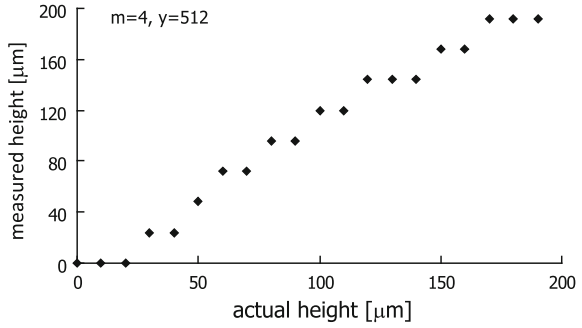
Z direction occur with the conventional light-section method, as well as the 3-D
shape measurement method using a multi-sided mirror described in this chapter.

6.5.2 Printed Circuit Board with Solders

Next, the 3-D shape of a 94 mm × 147 mmprinted circuit board of a mobile phone
was measured using the 3-D shape measurement system. The circuit board is shown
in Fig. 6.11. The circuit board had slightly oxidized solders, but no electrical
components were mounted on it. There were 3875 solders, and their height was
about 100–200 μm. The measurement period was 10 s.
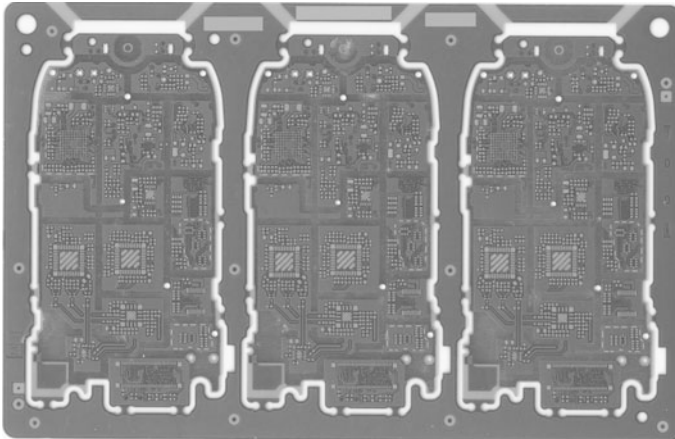
Fig. 6.11 Printed circuit board with solders

$t = 0.0$ s $t = 2.0$ s

$t = 4.0$ s $t = 6.0$ s

$t = 8.0$ s $t = 10.0$ s
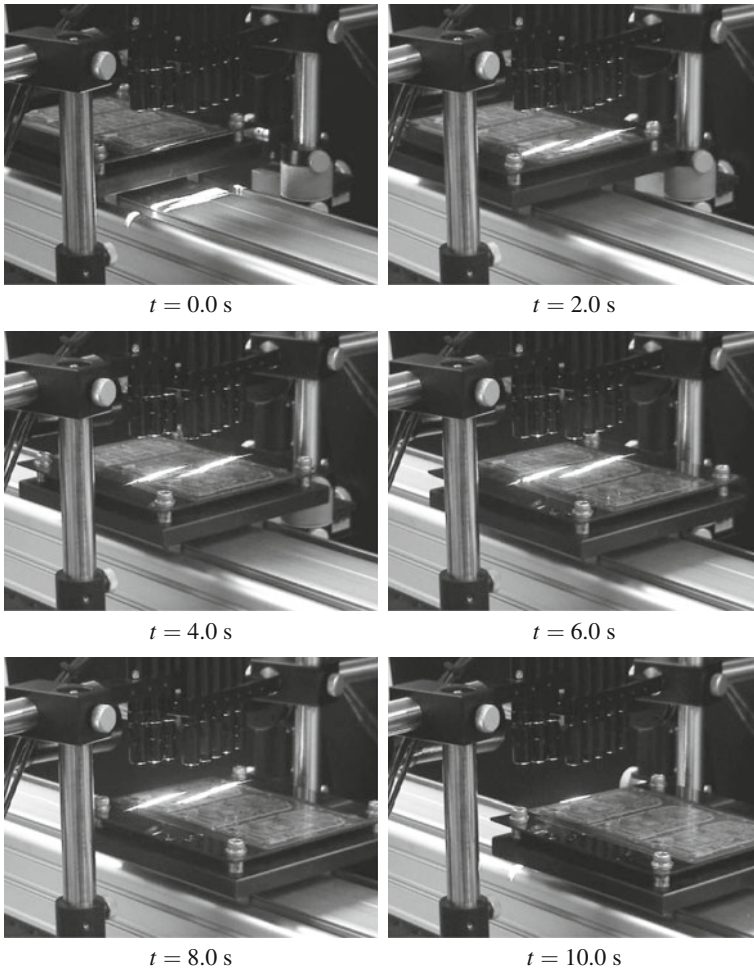
Fig. 6.12 Experimental images

Figure 6.12 shows the experimental scenes captured using a standard digital video camera at intervals of 2.0 s for $t = 0.0–10.0$ s. Figure 6.13 shows an integrated 3-D shape measurement result, which was measured using eight separate mirrors. The 3-D shape of the printed circuit board was measured during only one mechanical scan in the conveyance direction. Figure 6.14 shows the 3-D shape data for a local area of 14.4 mm × 14.4 mm, where the X range was 33.5–47.9 mm and the Y range was 43.1–57.5 mm. It can be seen that minute solders measuring 100 μm in height are visible in the 3-D shape graph.

Using the mask information from the CAD data of the printed circuit board being measured, the 3-D shapes were inspected for all 3875 solders by calculating their 3-D shape parameters, i.e., their volume, maximum height, and base area.

$^m P'_Y$ [mm]
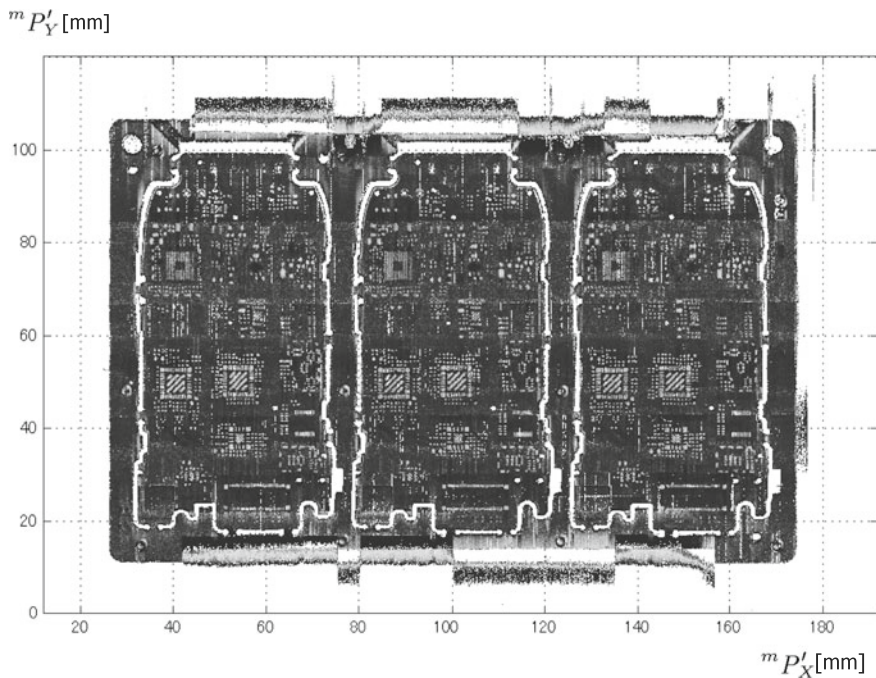
Fig. 6.13 3-D shape measurement result for an entire printed circuit board

Figure 6.15 shows the masked 3-D shapes of the solders, which correspond to the CAD mask data for a local area of 30 mm × 30 mm on the printed circuit board. In the experiment, all 3875 solders were inspected to determine their 3-D shape parameters and it was confirmed that the solders were covered with masks. Figure 6.16 shows images and the 3-D shapes measured for solders No. 97 and No. 81. Table 6.3 shows the 3-D shape parameters calculated based on their 3-D shape information. It can be seen that solder No. 97 was printed normally, whereas solder No. 81 was printed inadequately on the printed circuit board. Thus, the volume, maximum height, and base area of solder No. 81 were clearly smaller than those of solder No. 97 in Table 6.3. The experimental results showed that the newly developed HFR 3-D shape measurement system described in this chapter could quantify defective solders simultaneously on a printed circuit board using the 3-D shape parameters at the submillimeter level.

(a)

(b)
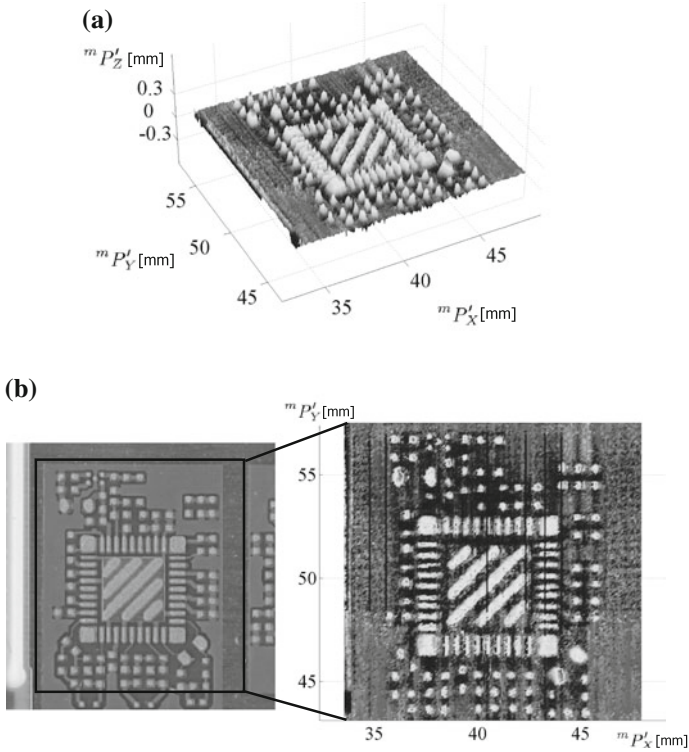
Fig. 6.14 3-D shape measurement results for a local area (14.4 mm × 14.4 mm): **a** calculated 3-D shape graph and **b** actual image and calculated 3-D shape

(a) **(b)**
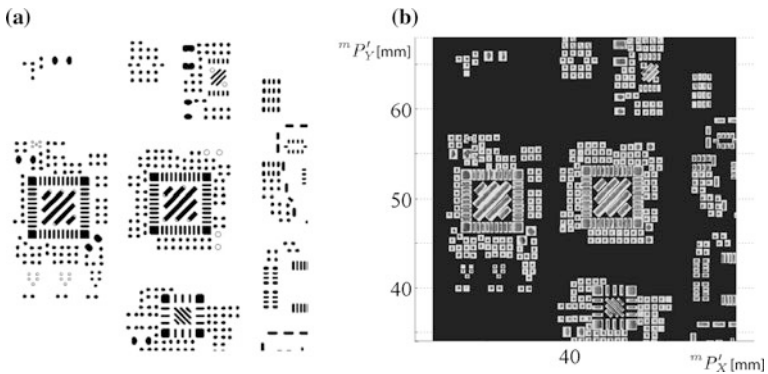
Fig. 6.15 Masking for 3-D shape inspection: **a** CAD mask data and **b** masked 3-D shapes of solders
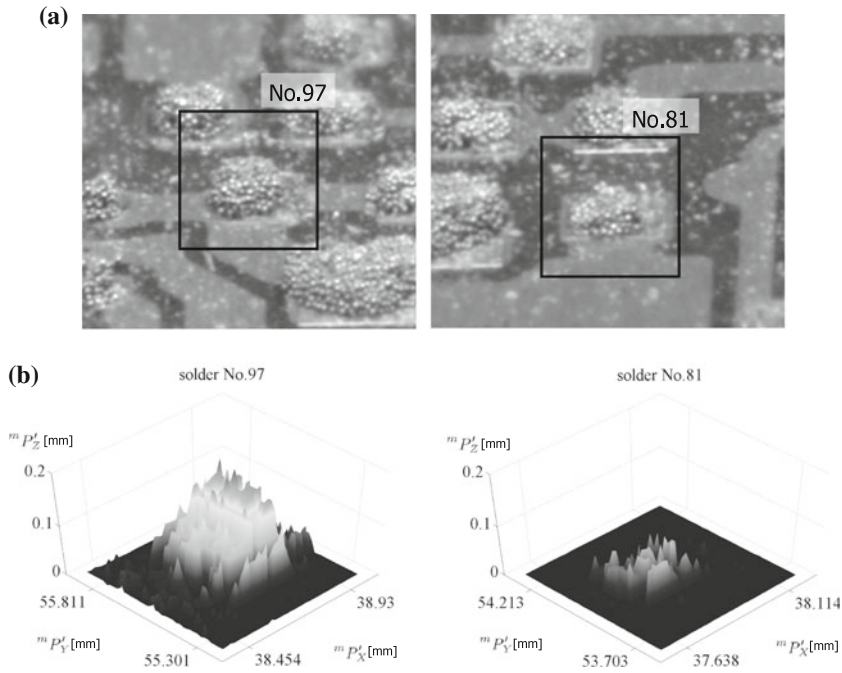
Fig. 6.16 3-D shapes of solders No. 97 and No. 81: **a** overviews of solders and **b** 3-D shapes

Table 6.3 3-D shape parameters of solders No. 97 and No. 81

Solder No. 97	
X-centroid [mm]	38.67
Y-centroid [mm]	55.59
Max-height [mm]	0.17
Volume [mm³]	0.012
Bottom area [mm²]	0.13
Solder No. 81	
X-centroid [mm]	37.82
Y-centroid [mm]	53.95
Max-height [mm]	0.08
Volume [mm³]	0.002
Bottom area [mm²]	0.05

6.6 Conclusion

This chapter describes a novel light-section method for 3-D shape measurement, which facilitates wider resolution measurements by an image sensor based on a spatial transform using a multi-sided mirror. The newly developed HFR 3-D shape measurement system uses a high-speed vision hardware system that allows real-time image processing at 1000 fps. The effectiveness of the 3-D shape measurement system was confirmed by several verification experiments using a printed circuit board with 3875 solders in a single mechanical scan, which included 3-D shape inspection using the CAD mask data.

References

1. Rioux M, Bechthold G, Taylor D, Duggan M (1987) Design of a large depth of view three dimensional camera for robot vision. Opt Eng 26(12):1245–1250
2. Blais F, Rioux M, Beraldin JA (1988) Practical considerations for a design of a high precision 3-D laser scanner system. Proc SPIE 959:225–246
3. Sun C, Shi H, Oiu Y, Ye S (2000) Line-structured laser scanning measurement system for BGA lead coplanarity. In: Proceedings of IEEE Asia-Pacific conference on circuits and systems, pp 715–718
4. Yen HN, Tsai DM (2004) A fast full-field 3D measurement system for BGA coplanarity inspection. Int J Adv Manuf Technol 24:132–139
5. Bernard TM, Zavidovique BY, Devos FJ (1993) A programmable artificial retina. IEEE J Solid-State Circuits 28(7):789–797
6. Komuro T, Kagami S, Ishikawa M (2004) A dynamically reconfigurable SIMD processor for a vision chip. IEEE J Solid-State Circuits 39(1):265–268
7. Ishii I, Yamamoto K, Kubozono M (2006) Higher order autocorrelation vision chip. IEEE Trans Electron Devices 53(8):1797–1804
8. Hirai S, Zakoji M, Masubuchi A, Tsuboi T (2005) Real-time FPGA-based vision system. J Robot Mechatron 17(4):401–409
9. Watanabe Y, Komuro T, Ishikawa M (2007) 955-fps real-time shape measurement of a moving/deforming object using high-speed vision for numerous-point analysis. In: Proceedings of IEEE international conference on robotics and automation, pp 3192–3197
10. Ishii I, Sukenobe R, Taniguchi, T., & Yamamoto, K. (2009). Development of high-speed and real-time vision platform, H^3 Vision. In: Proceedings of IEEE/RSJ international conference on intelligent robots and systems, pp 3671–3678
11. Ishii I, Tatebe T, Gu Q, Moriue Y, Takaki T, Tajima K (2010) 2000 fps real-time vision system with high-frame-rate video recording. In: Proceedings of IEEE international conference on robotics and automation, pp 1536–1541
12. Oike Y, Ikeda M, Asada K (2004) Design and implementation of real-time 3-D image sensor with 640 × 480 pixel resolution. IEEE J Solid-State Circuits 39(4):622–628
13. Oike Y, Ikeda M, Asada K (2005) A 375 × 365 high-speed 3-D range-finding image sensor using row-parallel search architecture and multi-sampling technique. IEEE J Solid-State Circuits 40(2):444–453
14. Lindgren L, Melander J, Johansson R, Moller B (2005) A multiresolution 100-GOPS 4-Gpixels/s programmable smart vision sensor for multisense imaging. IEEE J Solid-State Circuits 40(6):1350–1359

Chapter 7
Efficient Completeness Inspection Using Real-Time 3D Color Reconstruction with a Dual-Laser Triangulation System

Matteo Munaro, Edmond Wai Yan So, Stefano Tonello and Emanuele Menegatti

Abstract In this chapter, we present the final system resulting from the European Project "3DComplete" aimed at creating a low-cost and flexible quality inspection system capable of capturing 2.5D color data for completeness inspection. The system uses a single color camera to capture at the same time 3D data with laser triangulation and color texture with a special projector of a narrow line of white light, which are then combined into a color 2.5D model in real time. Many examples of completeness inspection tasks are reported, which are extremely difficult to analyze with the state-of-the-art 2D-based methods. Our system has been integrated into a real production environment, showing that completeness inspection incorporating 3D technology can be readily achieved in a short time at low costs.

Contents

M. Munaro (✉) · E.W.Y. So · E. Menegatti
Intelligent Autonomous Systems Laboratory, University of Padova,
via Gradenigo 6a, 35131 Padua, Italy
e-mail: munaro@dei.unipd.it

E.W.Y. So
e-mail: edmond.waiyan.so@gmail.com

E. Menegatti
e-mail: emg@dei.unipd.it

S. Tonello
IT+Robotics Srl, Strada Prima 11, 35129 Padua, Italy
e-mail: stefano.tonello@it-robotics.it

7.1 Introduction

An important application of machine vision is quality control and in particular completeness inspection—checking the presence or absence of parts, identifying their type, and determining their position and orientation. Until recently, many systems have relied on 2D-based image analysis [4]. However, without depth information, 2D machine vision systems have various shortcomings that limit their capabilities and robustness. For example, they cannot reliably detect parts in front of a background of similar color or reflectance (i.e., low contrast). Also, difficulties often arise when dealing with non-rigid objects, such as hoses and cables, that may change their position and lead to occlusions so that the parts to be inspected are not fully visible. As commercial 3D scanners become more affordable, the manufacturing industry is increasingly looking to incorporate 3D scanning technology to overcome the limitations of 2D machine vision [1]. The 3DComplete European project (http://3dcomplete.eu) is carried out by a consortium of small business enterprises and research institutes in Europe aiming to develop 3D completeness inspection systems that are similar to existing 2D inspection systems in terms of speed, ease of use, and integration, but provide improved functionality by incorporating 3D data. While commercial 3D sensing products already exist, they are designed mainly for reverse engineering and precise metrology and have prohibitively high costs. A major aim of the project is to create low-cost 3D sensing solutions that are suitable for completeness inspection.

This chapter describes one of the technologies being developed under this project, called *extended 2.5D inspection*, where 3D information obtained by laser triangulation is combined with color and texture information obtained with a camera. To demonstrate this technology, the University of Padova and IT +Robotics, two of the participants in the project, have developed a prototype color scanner using commercial-off-the-shelf components to showcase a low-cost sensor suite that can be readily deployed in an industrial assembly line setting to perform 3D completeness inspection.

A typical problem of laser triangulation systems is the trade-off between depth resolution and sensing range. Many commercial solutions, such as the SICK Ruler, have a fixed baseline. These products are already precisely calibrated, but have the drawback that they work only at a fixed resolution and sensing range. Moreover, the problem of occlusions cannot be mitigated by adding more laser projectors in different positions. Other solutions such as the SICK Ranger offer higher configurability by allowing range to be acquired from more than one laser, but the calibration between range and texture data in order to register the two types of information is left to the end user.

Our system, on top of being cost effective, provides the possibility to use two lasers for range acquisition and register the depth acquisition with the colored texture acquired in the same pass. Moreover, it offers full flexibility in the lasers position, thus allowing also asymmetric configurations that capture both the macroscopic geometry of the scanned object and minute surface details that may be important for defect inspection.

The reconstruction and inspection software has been developed using open-source libraries with permissive licenses. This demonstrates that in addition to inexpensive hardware, the software development cost for 2.5D technology can also be very low.

The remainder of the chapter is organized as follows: In Sect. 7.2, the main concepts related to laser triangulation systems are explained, while our dual-laser triangulation system is presented in Sect. 7.3. Section 7.4 details the calibration procedure we developed for our system, and Sect. 7.5 describes the reconstruction pipeline and products. In Sect. 7.6, we report some examples of completeness inspection we can perform with the output of our system, and an industrial solution is also presented in Sect. 7.7. Finally, conclusions are drawn in Sect. 7.8.

7.2 Laser Triangulation

The 2.5D color scanner being developed uses laser triangulation as the underlying technology to obtain 3D information. A laser triangulation system, also known variously as a laser striper and a laser stripe profiler, is a form of structured light system.

As shown in Fig. 7.1, such a system comprises of an illuminating laser source and a camera. Using the triangulation principle, with known baseline b and

Fig. 7.1 General
configuration of a laser
triangulation system
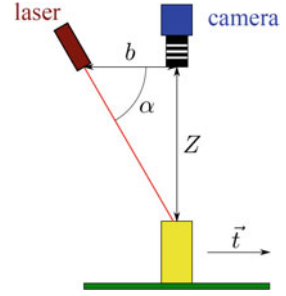
triangulation angle α between the laser and camera, the distance Z to the scanned object can be determined. The system parameters b and α affect several aspects of performance, such as the sensing range, depth resolution, occlusion, and laser reflectance, which we briefly describe in the following using 2D geometry. A more thorough analysis can be found in Pears [6].

7.2.1 Sensing Range

As shown in Fig. 7.2, for a laser located at $\mathcal{O}_L$, and a camera located at $\mathcal{O}_C$ with a field of view of ϕ, with baseline b between the laser and camera, and triangulation angle α between their principal axes, the minimum distance $Z_{\min}$ and maximum distance $Z_{\max}$ that can be measured by the system are as follows:

$$Z_{\min} = \frac{b \sin \alpha \cos \frac{\phi}{2}}{\cos\left(\alpha - \frac{\phi}{2}\right)} = \frac{b}{\frac{1}{\tan \alpha} + \tan\left(\frac{\phi}{2}\right)} \tag{7.1}$$

$$Z_{\max} = \frac{b \sin \alpha \cos \frac{\phi}{2}}{\cos\left(\alpha + \frac{\phi}{2}\right)} = \frac{b}{\frac{1}{\tan \alpha} - \tan\left(\frac{\phi}{2}\right)} \tag{7.2}$$

The total sensing range $\Delta Z = Z_{\max} - Z_{\min}$ is thus:

$$\Delta Z = b \sin \alpha \cos \frac{\phi}{2} \left(\frac{1}{\cos\left(\alpha + \frac{\phi}{2}\right)} - \frac{1}{\cos\left(\alpha - \frac{\phi}{2}\right)} \right) = \frac{2b \tan\left(\frac{\phi}{2}\right)}{\frac{1}{\tan^2 \alpha} - \tan^2\left(\frac{\phi}{2}\right)} \tag{7.3}$$

Thus, $Z_{\min}$, $Z_{\max}$, and ΔZ all increase with increasing baseline b and triangulation angle α.

Fig. 7.2 Sensing range of a laser triangulation system
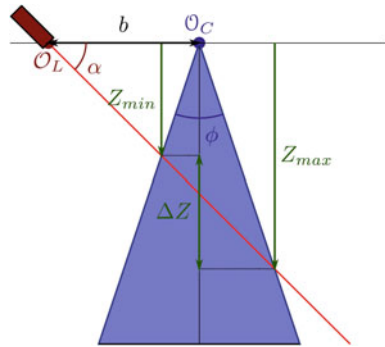
7.2.2   Depth Resolution

Using the pinhole camera model, for a camera with focal length f and pixel size sz_{px}, the angular resolution is $\theta = \tan^{-1} \frac{sz_{px}}{f}$. As shown in Fig. 7.3, with baseline b and triangulation angle α between the laser and camera, the depth resolution *res* at distance Z along the optical axis is as follows:

$$\text{res} = Z \left(\frac{\cos \alpha \cos \theta}{\cos(\alpha - \theta)} - 1 \right) = Z \frac{1}{\frac{1}{\tan \alpha \tan \theta} - 1} \tag{7.4}$$

Thus, the depth resolution res deteriorates with increasing triangulation angle α and at increasing depth Z.

Fig. 7.3 Depth resolution in a laser triangulation system
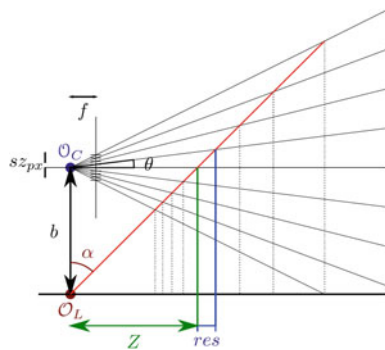
7.2.3 Occlusion

As shown in Fig. 7.4, in a laser triangulation system, occlusion can occur either when the laser light is blocked by an object from reaching another surface or when the camera is blocked by an object from imaging the laser light on another surface. We will refer to them respectively as projection path occlusion and imaging path occlusion.

We can quantify occlusion by measuring the distance on a horizontal surface lying at distance Z_2 that would be occluded by an object at distance Z_1. For a given baseline b and triangulation angle α, the amount of projection path occlusion and image path occlusion are respectively as follows:

$$\mathrm{occ}_{\mathrm{prj}} = \frac{Z_2 - Z_1}{\tan \alpha} \tag{7.5}$$

$$\mathrm{occ}_{\mathrm{img}} = \left(1 - \frac{Z_2}{Z_1}\right)\left(b - \frac{Z_2}{\tan \alpha}\right) \tag{7.6}$$

Thus, with increasing triangulation angle α, projection path occlusion $\mathrm{occ}_{\mathrm{prj}}$ decreases, while imaging path occlusion $\mathrm{occ}_{\mathrm{img}}$ increases.

7.2.4 Laser Reflectance

The amount of light reflected at point P on a surface is known as the scene radiance L. Using the Lambertian reflectance model, as shown in Fig. 7.5a, under illumination by a source at direction $\hat{I} = I/\|I\|$ and intensity $\|I\|$, the scene radiance
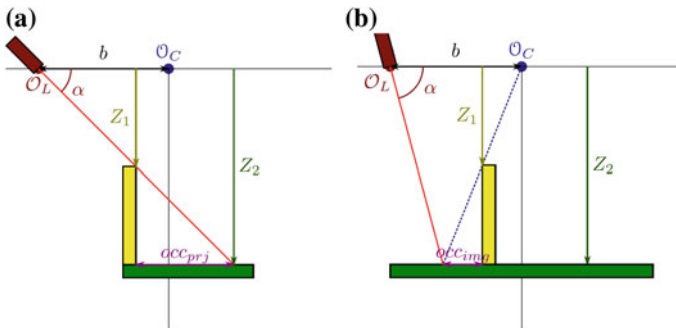
Fig. 7.4 Occlusion in a laser triangulation system. **a** Projection path occlusion. **b** Imaging path occlusion
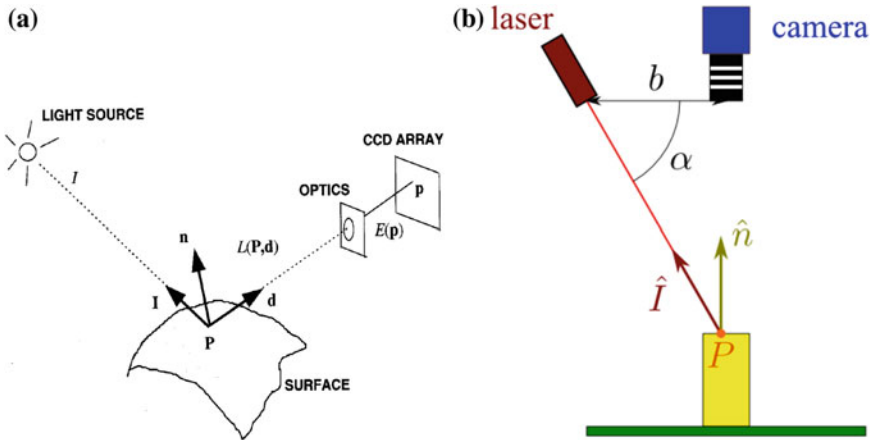
Fig. 7.5 Laser reflectance in a laser triangulation system. **a** Radiometry under a Lambertian reflectance model [11]. **b** Scene radiance for a horizontal surface

$L(P)$ of a point P on a diffuse or matte surface with surface normal $\hat{n}$ and surface albedo ρ is [11]:

$$L = \rho I^T \hat{n} = \rho \|I\| \hat{I}^T \hat{n} \qquad (7.7)$$

As shown in Fig. 7.5b, for a horizontal surface, $\hat{I}^T \hat{n} = \sin \alpha$, and $L_{\text{horizontal}}(P) = \rho \|I\| \sin \alpha$. Thus, scene radiance increases with increasing triangulation angle α.

The actual amount of light imaged by a camera, known as the image irradiance E, and measured as the power of light received per unit area at a point p on the image plane, is [11]:

$$E(p) = L(P) \frac{\pi}{4} \left(\frac{D}{f}\right)^2 \cos^4 \Theta \qquad (8)$$

Here, Θ is the angle between the camera's optical axis and the ray from p to P, and D is the diameter of the lens. Thus, for a given scene radiance L, the image irradiance E decreases for pixels lying further from the principal image point $[u_c, v_c]$.

7.3 Dual-Laser Triangulation System

The 2.5D color scanner that was developed uses laser triangulation as the underlying technology to obtain 3D information. This is augmented with color texture information from a color CCD camera. As we will describe in Sect. 7.5, the depth,

Fig. 7.6 Dual-laser
triangulation system
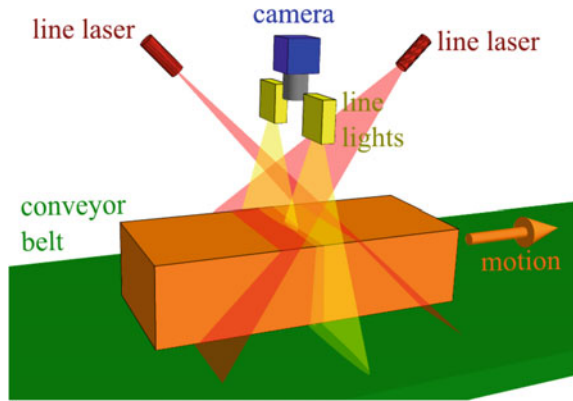
texture, and color information can be used either individually or in combination to produce a variety of data representations. This leads to a high degree of flexibility and robustness in creating inspection algorithms according to the task to be performed. Figure 7.6 shows the layout of the color 2.5D scanner that was developed, which we call a dual-laser triangulation system (DLTS). It consists of a color CCD camera, two line lasers, and a pair of line lights mounted over a translational platform. The major features of this configuration are described below.

7.3.1 Dual Laser for Reducing Occlusions

In designing our system for completeness inspection, the main focus was on being able to reconstruct large overall features of a scanned object. As a result, an important factor was to minimize the amount of occlusions in order to detect all the components on an object reliably. Thus, two line lasers are used in our configuration. Although a system with two cameras and one line laser [10, 12] can also be used to reduce occlusions, a dual-laser configuration is more cost effective, as an additional camera is usually more expensive than an additional line laser.

Figure 7.7a shows an example product to be inspected. Figure 7.7b shows a point cloud reconstruction of this object obtained with a single-pass scanning, with the points reconstructed by the two lasers differentiated by color. In particular, opposing vertical faces are often only detectable by one of the two lasers, as shown in Fig. 7.7b. A white connector highlighted in red in the images is of particular interest for inspection. It is located close to the base and surrounded by many other components; while it would be occluded to one of the lasers, a dual-laser configuration allows the component to be reliably detected regardless of the scanning direction.
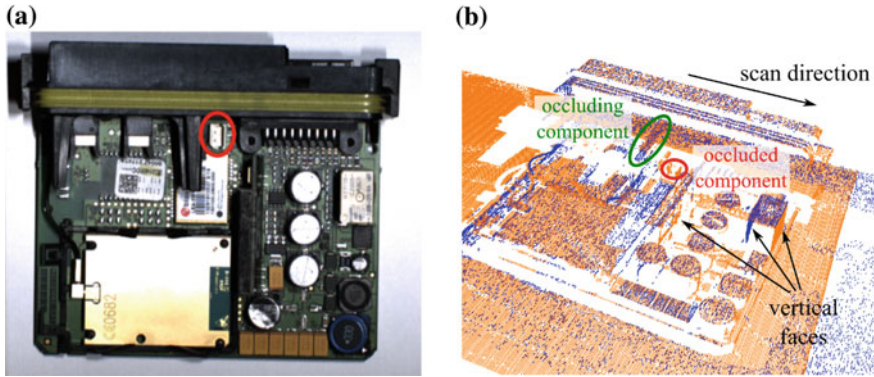
Fig. 7.7 Reduction of occlusions with dual lasers. **a** Color texture of example product, with white connector to be inspected highlighted in *red*. **b** Point cloud obtained with a single-pass scanning, with points reconstructed by the two lasers differentiated by color; the white connector is occluded to one of the lasers but visible to the other

7.3.2   Line Light for Simultaneous Color Texture Acquisition

Two line lights positioned on either side of the camera and aligned in the same direction of the laser lines (Fig. 7.8) are used to illuminate precisely the central part of the imaged scene, so that color texture information of the scanned object can be acquired simultaneously with the images of the laser lines, as shown in Fig. 7.9.

The line light consists of a series of low-power LEDs, and the illumination is focused by a cylindrical lens to produce the required beam width. For our operations, the beam is focused to a width of 10 mm at an operating distance of ∼ 300 mm. This can be easily modified by changing the relative distance between the LEDs and the lens for small adjustments, and lens diameter for larger adjustments. The small dimensions of the line lights enable them to be easily integrated with the sensor assembly and can be produced very economically. This solution was designed and produced for us by LightCube srl (http://www.light-cube.com).

Fig. 7.8 Line light assembly
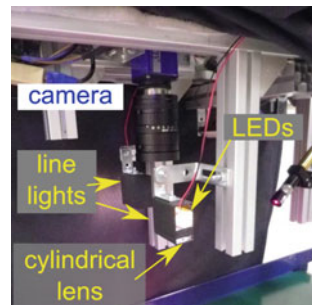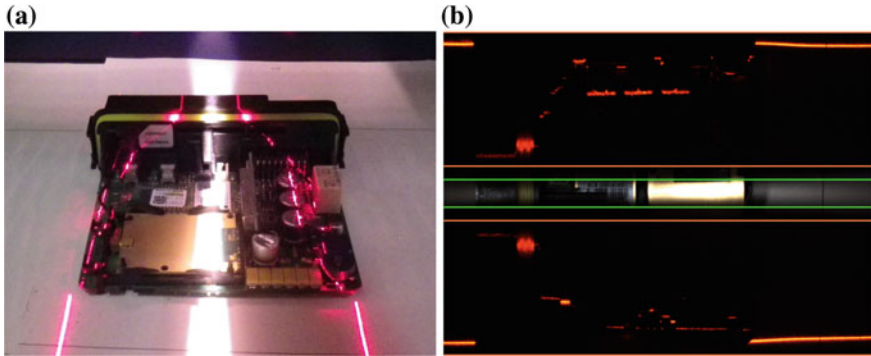
Fig. 7.9 Simultaneous laser lines and color texture acquisition. **a** Illumination of object by line lights and line lasers. **b** The imaged scene is separated into three regions for simultaneous laser lines and texture acquisition

By acquiring laser lines and texture together in the same image, the number of image rows available for the laser line detection is decreased, leading to a reduction in the depth resolution that would otherwise be achievable. However, when using a dual-laser configuration, a small separation should be reserved between the image regions of the two lasers in order to avoid the laser lines crossing into the other region. Furthermore, only very small number of image rows (5–10 in our operations) is required for obtaining a continuous texture image. Thus, in our implementation, there is practically no impact on the maximum achievable depth resolution.

An alternative approach for acquiring texture information would be to operate the lasers and an area lighting system in an alternating sequence, so laser line and color images are acquired separately in succession. However, this would require additional hardware for synchronization, and the frame rate of the camera would have to be divided between the two types of images, reducing reconstruction density.

7.3.3 High Configurability

Another advantage of using a dual-laser configuration is that the lasers can be positioned asymmetrically to achieve different sensing range and depth resolutions (Fig. 7.10). In general, the trade-off with using a smaller triangulation angle α to achieve a higher depth resolution *res* is a reduction in the sensing range ΔZ, as well as an increase in occlusions (Fig. 7.10a). However, in a dual-laser configuration, the two lasers can effectively complement each other in terms of these opposing factors. Thus, one laser can be used to capture the macroscopic geometry of the scanned
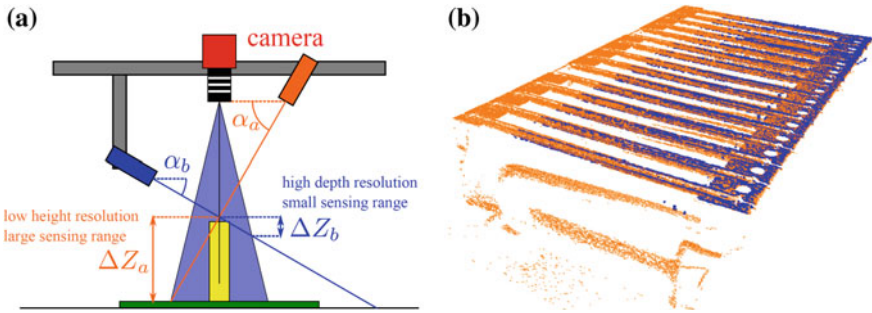
Fig. 7.10 Asymmetric dual lasers configuration. **a** Two lasers can be configured for different depth resolution and sensing range. **b** This allows one laser to be used for obtaining the macroscopic geometry, while the other is used to obtain minute surface details

object, while the other laser can be used to focus on minute surface details that may be important for defect inspection (Fig. 7.10b).

7.3.4 Hardware Components and System Prototype

Our DLTS sensor suite is assembled using readily available commercial machine vision components. The chosen color CCD camera has a resolution of 1600 × 1200, with a full-frame frame rate of 28.1 fps. It is paired with a 16-mm lens. The 5-mW red lasers emit at a wavelength of 635 nm, and its accompanying focusable optics are used to create a Gaussian line profile.

To test the performance of the DLTS in a realistic assembly line environment, a system prototype was built by mounting the DLTS sensor suite over a conveyor belt (Fig. 7.11). The system is enclosed using dark curtains to prevent ambient lighting from interfering with image acquisition. However, in an industrial setting, the need for curtains can be avoided by using more powerful lasers.

The lasers, line lights, and conveyor belt are controlled by a PC workstation through a National Instruments USB-6009 digital I/O unit; this function would

Fig. 7.11 System prototype with DLTS mounted over conveyor belt
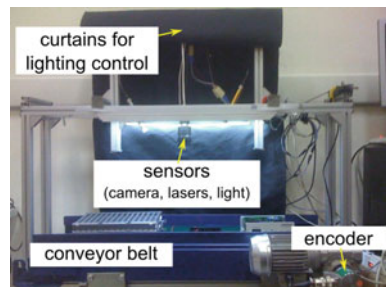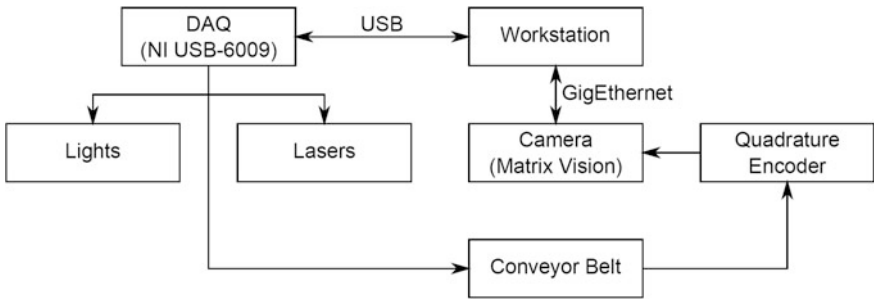
Fig. 7.12 Schematic of hardware components

normally be replaced by a PLC in an industrial environment. The encoder signals are directly connected into the camera, so that the conveyor belt position is precisely recorded with each frame at the time of acquisition. The camera transmits image and encoder data to the workstation through a Gigabit Ethernet link (Fig. 7.12).

The total cost of building the 2.5D color scanner is $\sim$€3000. The conveyor belt, encoder, and digital I/O device are excluded, as they are assumed to be a part of the assembly line into which the system is integrated. Because the prices considered are the unit-piece purchase costs, the final cost can be expected to be reduced. Our system is 2–3 times less expensive than existing commercial sensors offering similar capability and performance.

7.3.5 Reconstruction Density and Resolution

The primary output of the 2.5D color scanner is a reconstruction of the scanned object in the form of a color point cloud, as shown in Fig. 7.13b. The density and resolution of this reconstruction is determined by a number of system parameters, including the camera frame rate (fps), the speed of the conveyor belt (v), the
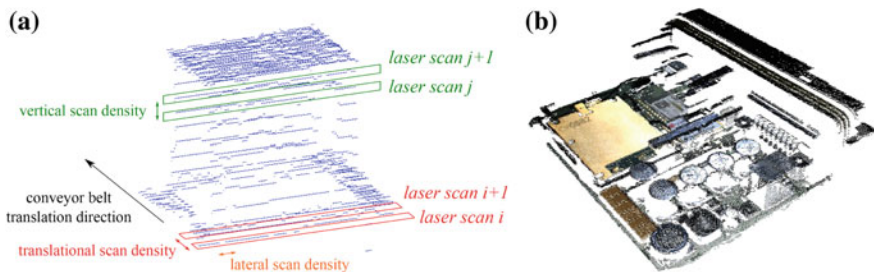
Fig. 7.13 **a** Reconstruction density of color point cloud. **b** Color point cloud reconstructed using our 2.5D color scanner

Table 7.1 Reconstruction density and resolution

Parameter	Formula	Value (mm)
Translation scan density	$\frac{v}{\text{fps}}$	0.534
Vertical scan density	$\frac{v}{\text{fps}}\tan\alpha$	1.145
Lateral scan density	$Z\frac{sz_{px}}{f}$	0.096
Vertical resolution	$Z\frac{sz_{px}\tan\alpha}{f-sz_{px}\tan\alpha}$	0.207

triangulation angle (α), and the camera to scene distance (Z), all of which can be adjusted to adapt to different inspection requirements. Figure 7.13a illustrates the terms we use to characterize our reconstructed 3D models, while Table 7.1 gives the formulas for these characterizations, as well as the values for our current setup, with fps = 28, v = 15 mm/s, α = 65°, Z = 350 mm, f = 16 mm, and sz_{px} = 4.4 µm. This reconstruction density and resolution are adequate for most of the completeness inspection tasks that we have considered, ranging from a small PCB board measuring 9 × 12 cm and 3 cm high, to larger automotive components and heating elements measuring 20 × 30 cm and 10 cm high.

7.4 Sensor Calibration

The calibration of the system involves the recovery of three sets of parameters (Fig. 7.14):

- camera intrinsic parameters (focal length f, principal point $[u_c, v_c]^T$, distortion coefficients)
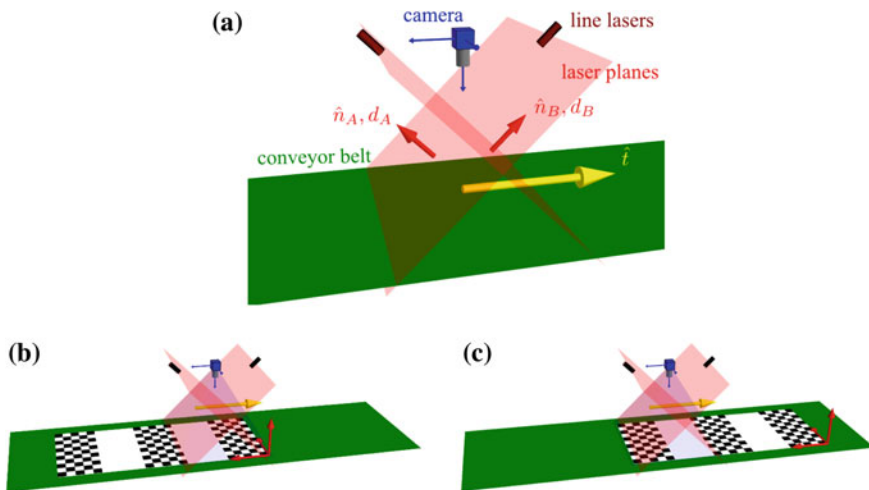
Fig. 7.14 a Sensor calibration parameters. **b–c** Calibration procedure using a planar checkerboard

- laser light plane parameters $(\hat{\mathbf{n}}_A, d_A, \hat{\mathbf{n}}_B, d_B)$
- translational direction of the conveyor belt $(\hat{\mathbf{t}})$.

We refer to the recovery of these sets of parameters respectively as camera calibration, laser calibration, and conveyor calibration. McIvor [5] has given a holistic calibration procedure for a similar system that recovers all three sets of parameters simultaneously, but requires the use of a 3D calibration object. For our system, we have devised a simple procedure requiring only a planar checkerboard pattern (Fig. 7.14b, c). In addition to being easily and inexpensively produced, the checkerboard pattern can be readily detected using open-source software such as OpenCV, reducing the software development costs.

7.4.1 Misalignment Between Point Clouds

The calibration method we devised is similar to mainly existing techniques for camera-laser calibration [2, 3, 7, 13, 14]. However, for our dual-laser configuration, small errors in the calibration can result in a misalignment between the point clouds reconstructed with the two lasers, arising from an inconsistency between the laser plane parameters $(\hat{\mathbf{n}}_A, d_A, \hat{\mathbf{n}}_B, d_B)$ and the conveyor motion $(\hat{\mathbf{t}})$ when they are determined independently. As shown in Fig. 7.15, a reconstruction of the conveyor belt plane and a rectangular block reveals a significant offset between the point clouds of the two lasers. The offset in the vertical direction ranges from 3 to 5 mm.

Figure 7.16 illustrates for the 2D case how this misalignment depends on errors in calibration data. When laser calibration and motion calibration are performed separately, the misalignment cannot be detected from the calibration results.
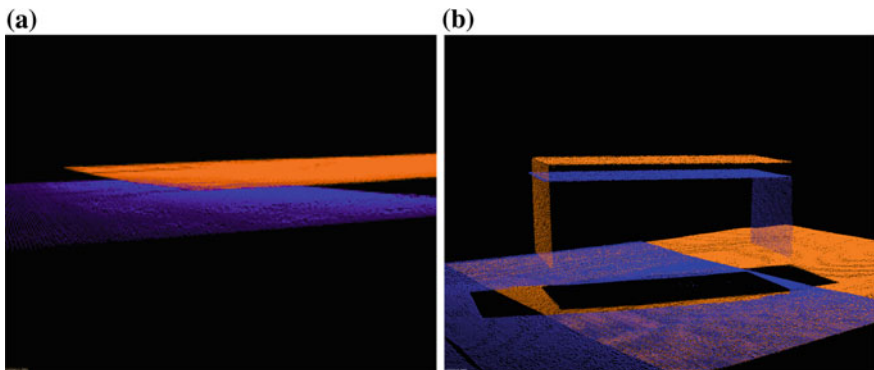
Fig. 7.15 Reconstructions showing offset between the point clouds of the two lasers. **a** Conveyor plane. **b** Rectangular block
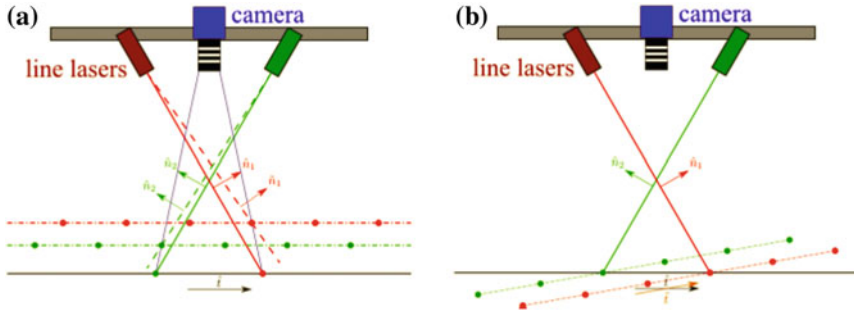
Fig. 7.16 Misalignment between the two laser point clouds in the reconstruction of a single plane due to (**a**) errors in laser plane parameters and (**b**) errors in translation direction

7.4.2 Constrained Calibration

In order to detect and correct the misalignment between the point clouds during calibration, a calibration procedure in which the calibration data are constrained by both the laser and motion parameters is required. We present two solutions: The first method uses the laser plane parameters to constrain the motion calibration, while the second method uses the motion parameters to constrain the laser calibration.

7.4.2.1 Constraining Motion Calibration with Laser Parameters

First, laser calibration is performed independently from motion calibration by using the method described in So et al. [9]. Next, motion calibration is performed as follows. A calibration board with three checkerboard patterns is placed parallel to the conveyor belt as illustrated in Fig. 7.14b, c. The line lasers are also turned on. Images of the checkerboard patterns and the laser lines are captured together at various positions, along with either the encoder counts ΔE_i or the timestamp Δt_i.

Using the laser plane parameters alone, the detected laser points are triangulated, giving their 3D coordinates with respect to the camera center. Because two line lasers are present, the laser range data can be used to determine the plane π_c of the calibration board. The plane parameters $\pi_c = (\hat{\mathbf{n}}_c, d_c)$ are recovered by finding the best-fit plane to all of the triangulated laser points. To recover the translational direction $\hat{\mathbf{t}}$ of the conveyor belt, we form a constrained pose estimation problem with the following constraints:

- because the calibration board is undergoing pure translation, all of the poses have the same rotation, $\mathbf{R} = \mathbf{R}_1 = \cdots = \mathbf{R}_i$
- the translation $\overrightarrow{\mathbf{t}_i}$ from the camera to each checkerboard pose is offset from the first checkerboard pose $\overrightarrow{\mathbf{t}_1}$ in the same direction $\hat{\mathbf{t}}$, $\overrightarrow{\mathbf{t}_i} = \overrightarrow{\mathbf{t}_1} + \mu_i \hat{\mathbf{t}}$

- all of the checkerboard poses $\left(\mathbf{R}_i, \vec{\mathbf{t}}_i \right)$ must lie on the calibration plane $\pi_\mathbf{c}$ defined by the laser points

where μ_i is the magnitude of the translation motion, and $\hat{\mathbf{t}}$ is the translational direction being seek in the current motion calibration.

The third constraint is the additional contribution of using lasers in motion calibration. To perform this constrained calibration, we establish an arbitrary calibration pose $\left(\mathbf{R}_c, \vec{\mathbf{t}}_c \right)$ on the plane $\pi_\mathbf{c}$ and minimize the following cost function:

$$\arg \min_{\psi, \, \vec{\mathbf{t}}_{c1}, \theta, v} \sum_j \left\| \mathbf{x}_j - \mathbf{K} \cdot \left[\mathbf{R} | \vec{\mathbf{t}}_i \right] \cdot \mathbf{X}_j \right\|^2 \tag{9}$$

Here, $\mathbf{X}_j$ and $\mathbf{x}_j$ are the checkerboard corners and their imaged coordinates, $\mathbf{K}$ is the camera intrinsic parameters matrix, v is the speed of the linear motion platform, ψ is the angle of rotation of the checkerboard patterns from the calibration pose $\mathbf{R} = \mathbf{R}_c \cdot \mathbf{R}_z(\psi)$, and $\vec{\mathbf{t}}_{c1}$ is a 2D vector lying in the calibration plane $\pi_\mathbf{c}$ corresponding to the offset of the first checkerboard pose from the calibration pose, $\vec{\mathbf{t}}_1 = \vec{\mathbf{t}}_c + \mathbf{R}_c \cdot \left[\vec{\mathbf{t}}_{c1}^T, 0 \right]^T$. Moreover, when the calibration board is placed parallel on the conveyor belt, $\hat{\mathbf{t}}$ must lie in the plane of the calibration board. This is parameterized using 1 DOF, $\hat{\mathbf{t}} = \cos \theta \hat{\mathbf{i}} + \sin \theta \hat{\mathbf{j}}$, where $\hat{\mathbf{i}} = \mathbf{R} \cdot [1 \quad 0 \quad 0]^T$ and $\hat{\mathbf{j}} = \mathbf{R} \cdot [0 \quad 1 \quad 0]^T$. Thus, the constrained problem has 4–5 DOF, corresponding to 3 DOF for the first checkerboard pose, 1 DOF for $\hat{\mathbf{t}}$, and 1 DOF for v if desired.

7.4.2.2 Constraining Laser Calibration with Motion Parameters

First, motion calibration is performed independently from laser calibration by using the method described in So et al. [9]. However, the laser lines are imaged together with the checkerboard patterns, and the process is repeated multiple times with the calibration board at different orientations $\mathbf{R}^k$ (Fig. 7.17).
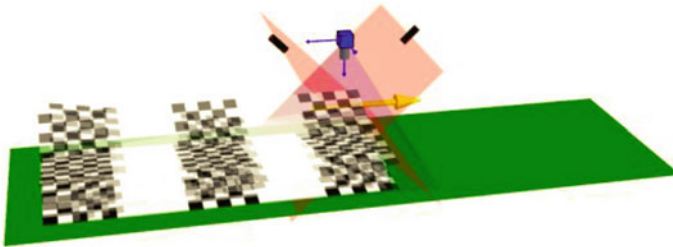
Fig. 7.17 Constrained laser calibration

Using the correspondences $\left\{ \mathbf{X}_j^k \leftrightarrow \mathbf{x}_j^k \right\}$ between the checkerboard corners and their imaged coordinates, a single translational direction $\hat{\mathbf{t}}$ is estimated for the entire data set by minimizing the following cost function:

$$\arg \min_{\mathbf{R}^k, \overrightarrow{\mathbf{t}}^k_{1,\theta,v}} \sum_j \left\| \mathbf{x}_j^k - \mathbf{K} \cdot \left[\mathbf{R}^k | \overrightarrow{\mathbf{t}}_i \right] \cdot \mathbf{X}_j^k \right\|^2 \tag{10}$$

Here, $\left(\mathbf{R}^k, \overrightarrow{\mathbf{t}}^k_1 \right)$ is the initial pose of the calibration board for orientation $\mathbf{R}^k$. We assume that one of the orientations $\mathbf{R}^p$ is parallel to conveyor belt, so that the translational direction $\hat{\mathbf{t}}$ can still be parameterized using 1 DOF, $\hat{\mathbf{t}} = \cos \theta \hat{\mathbf{i}} + \sin \theta \hat{\mathbf{j}}$, where $\hat{\mathbf{i}} = \mathbf{R}^p \cdot [1 \quad 0 \quad 0]^T$ and $\hat{\mathbf{j}} = \mathbf{R}^p \cdot [0 \quad 1 \quad 0]^T$. For n_k different checkerboard orientations, the constrained problem has $6n_k + 1$ or $6n_k + 2$ DOF.

Using the checkerboard poses $\left(\mathbf{R}^k, \overrightarrow{\mathbf{t}}^k_i \right)$ and the laser points in the same images, laser calibration can be performed using the standard method described in So et al. [9].

7.4.2.3 Validation of Constrained Calibration

To validate the constrained calibration methods presented above, a plane is placed at ten different positions and scanned, and the offset between the point clouds reconstructed from the two lasers is measured. As shown in Fig. 7.18, when laser calibration and motion calibration are performed independently, the misalignment error ranges between 0.5 and 5 mm. When the motion calibration or the laser calibration is constrained by the other set of parameters, the error is dramatically reduced to 1.0 and 0.4 mm, respectively.

Fig. 7.18 Misalignment errors between the point clouds of the two lasers of a plane scanned at 10 different positions
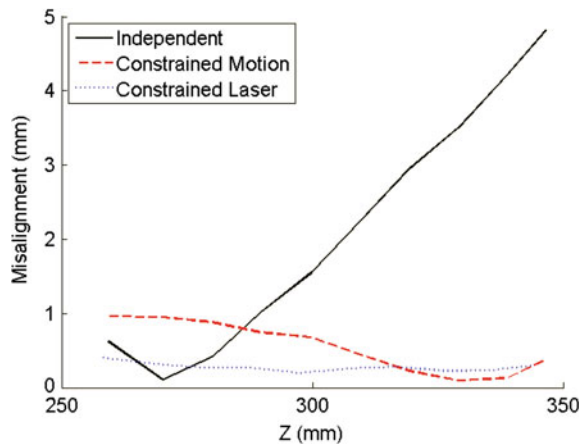
7.5 3D Reconstruction

7.5.1 *Reconstruction Process and Output Products*

Figure 7.19 shows the pipeline of the reconstruction software developed for our dual-laser triangulation system with a color camera, along with the different data products it produces. The laser points are detected in the images, triangulated and merged together by means of calibration to obtain a 3D point cloud of the scanned object, which can be used to derive a range image. On the other hand, the central part of each image contains texture which is used for creating a stitched image of the object or for texturing the 3D point cloud in order to obtain a direct correspondence between 3D and color information. The registered image is obtained superimposing the range image onto the texture image. Finally, a textured mesh can also be created from the colored point cloud for better visualization.

The different data representations can be exploited for different types of inspection tasks. For a detailed description on how these products are obtained, please refer to [8].
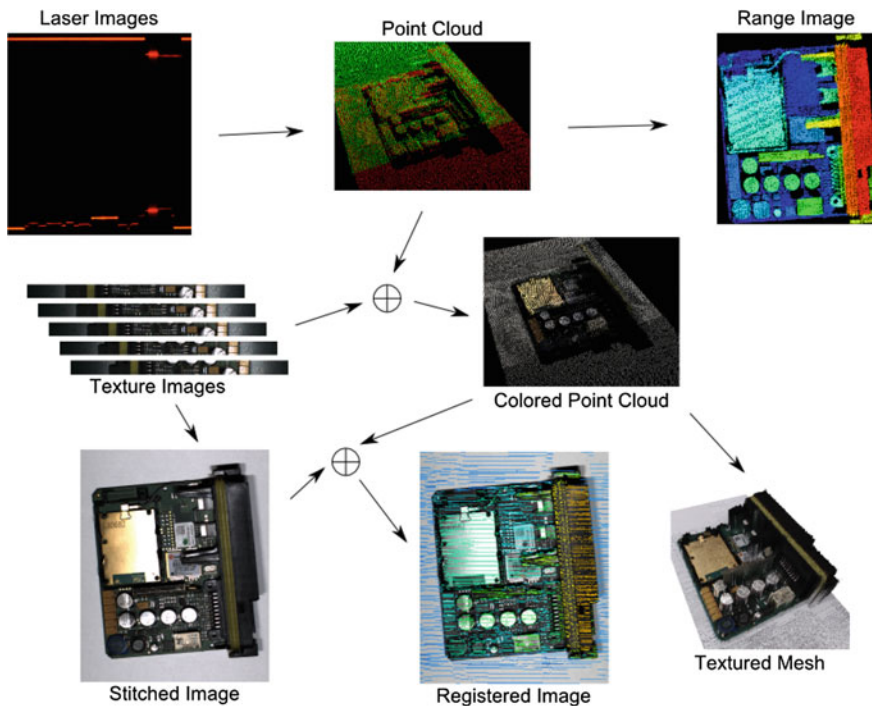
Fig. 7.19 Reconstruction pipeline and products

Table 7.2 Processing time for reconstruction

Process	Time (ms)
Laser detection and triangulation	16
Point cloud coloring	1
Texture stitching	5
Total	22

7.5.2 Software Performance

Table 7.2 shows the average processing time per frame required for the 3D reconstruction steps, executed on a standard workstation (Intel Xeon E3-1225 quad-core @ 3.1 GHz, 4 GB memory). At a frame rate of 28.1 fps, the inter-frame interval is 35.6 ms. Thus, the reconstruction pipeline can perform in real time.

The range image is generated after the entire point cloud has been constructed and can be completed extremely quickly in the order of milliseconds. The registered texture image is generated as part of the point cloud coloring and image stitching processes. Therefore, performance time is not reported for these processes.

7.6 Completeness Inspection

In this section, we report a number of examples of inspection tasks that can be solved by exploiting the products obtained in real time with the dual-laser triangulation system described in the previous sections. While many of the algorithms are trivial, it exemplifies the ease with which a wide range of inspection tasks can be accomplished when aligned 3D, and color information is available. The variety of data representations that is produced by our reconstruction software brings further flexibility in devising the inspection algorithms. In particular, we show how the color point cloud, the range image, the stitched texture image, and the registered texture image can be used effectively to solve different types of inspection tasks.

In some cases, quality inspection is obtained by aligning the scan of the object to inspect a reference model which does not present defects and then by searching for components of interest at known positions or by comparing the models in a direct manner, as we will see in Sect. 7.6.5. For aligning new scans with a reference model, we exploit the pose estimation and model alignment method described in So [8], which is based on principal component analysis and alignment of plane clusters.

7.6.1 Inspection Using Color Point Cloud

As an example of complex inspection task which can be solved with the high-resolution color point clouds, we report the cable inspection in Fig. 7.20. The
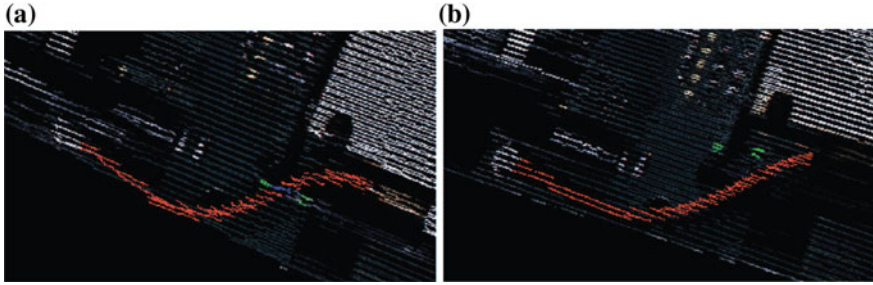
Fig. 7.20 Cable routing inspection using color point cloud. **a** Cable (*red*) correctly routed through pegs (*green*). **b** Incorrect cable routing

task was to detect whether a black cable with a 1 mm diameter was correctly routed between two vertical black pegs 4 mm high. This inspection task is very difficult to solve for traditional 2D visual inspection systems, because of the low contrast between the components. Instead, from our 3D point clouds, we can easily isolate the pegs and the cable. Our approach aims at identifying clusters which are not only coherent in position and number of points, but also have a particular curved shape. For this purpose, we select only points with mean curvature above 0.01. The estimated cable points are highlighted in red, while the pegs are colored in green. It can be noticed that both the cable and the pegs are correctly detected in both the good and the faulty boards.

7.6.2 Inspection Using Range Image

Another common inspection task is to check whether a component is present or absent. Figure 7.21a shows a sample PCB where the fuses circled in red are to be
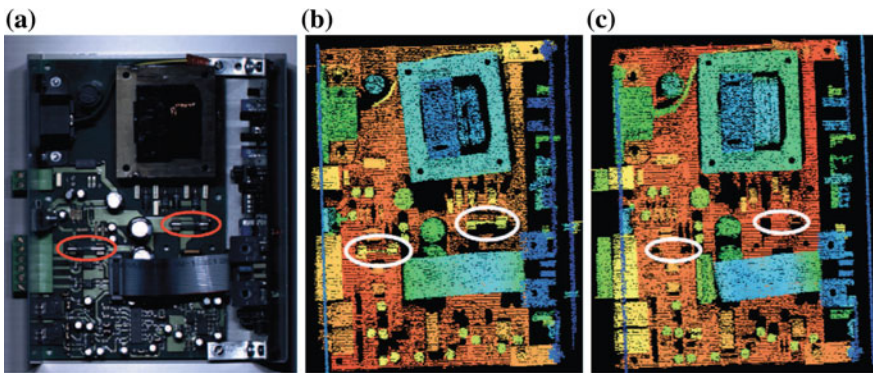
Fig. 7.21 Examples of inspection using range image. **a** Texture image showing location of clear fuses. **b** Range image showing the presence of fuses. **c** Range image showing the absence of fuses

checked. This is particularly difficult using the image alone, because the fuse is mostly a glass casing with a fine filament inside, and the ends of the fuse are metal caps that are of the same color and shape as their clips. Using the point cloud also proved to be difficult, because the sparse amount of points reflected off the glass does not allow them to be clustered together easily to be identified as a component. However, the range image in Fig. 7.21b gives a clear indication when they are present (component shows up as yellow, circled in white), and the range image in Fig. 7.21c gives a clear indication when they are absent.

7.6.3 Inspection Using Stitched Image

Another common inspection task is to check for the orientation of a component. In Fig. 7.22, the white two-pin connector on the right side of the image is to be checked, with Fig. 7.22a showing the correct orientation and Fig. 7.22d showing the incorrect orientation. Because the component is fairly small, the point cloud density is not high enough for the inspection to be made reliably using the point cloud (Fig. 7.22b, e) or the range image (Fig. 7.22c, f). However, the dense texture
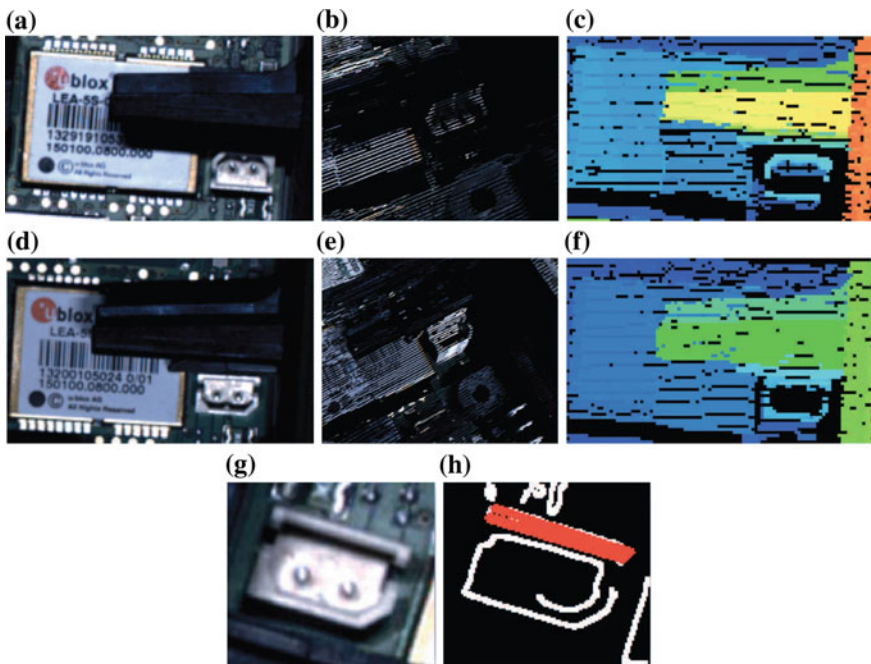
Fig. 7.22 Examples of inspection using texture image. **a** Correct texture image. **b** Correct point cloud. **c** Correct range image. **d** Incorrect texture image. **e** Incorrect point cloud. **f** Incorrect range image. **g** Detail of the texture image. **h** Edge and contour detection on texture image

information allows this check to be accomplished using 2D image analysis by edge
and contour detection, as shown in Fig. 7.22h.

7.6.4 Inspection Using Registered Image

A rather unique inspection task requires that the number of loops in a coil to be
counted, as shown in Fig. 7.23a.

The point cloud density is not high enough for each coil to be reconstructed
consistently from end to end, as seen in the point cloud (Fig. 7.23b) and the range
image (Fig. 7.23c). Inspection of the stitched texture image (Fig. 7.23d) shows that a
possible solution is to count the number of saturated pixels; however, saturated pixels
can belong to the top or a bottom of a loop, and it is difficult to differentiate between the
two using the texture image alone. The registered texture image (Fig. 7.23f) provides
an elegant solution here, since the depth of the saturated pixels can be determined,
allowing them to be classified reliably as either the top or bottom of a loop.

Given that only registered depth and texture data allow solving this problem, we
showed a clear advantage of our system with respect to many more expensive
commercial solutions which do not provide this kind of alignment.

7.6.5 Automatic Change Detection

Given that our system is highly repeatable in producing scans of an object and a new
scan can be aligned to a reference point cloud, a direct method can also be applied to
detect changes in a test point cloud with respect to a reference model. Figure 7.24a is
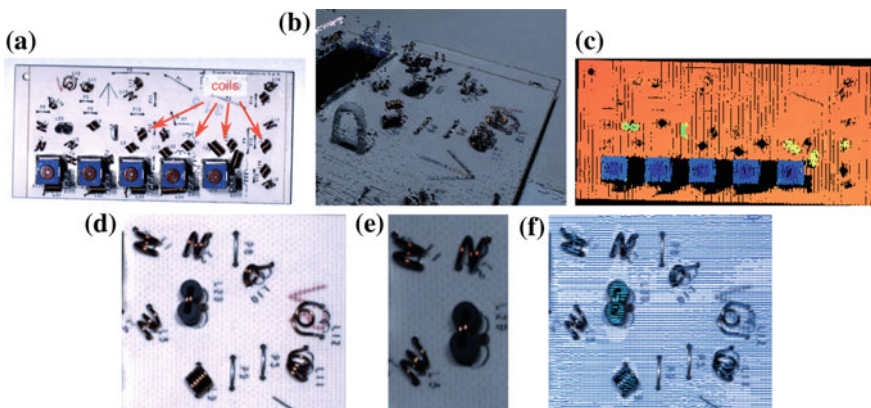
Fig. 7.23 Example of inspection using registered texture image. **a** Image showing location of
coils to be inspected. **b** Color point cloud. **c** Range image. **d** Texture image. **e** Saturated pixels in
texture image corresponding to coil elements. **f** Registered texture image used to determine the
depth of saturated pixels
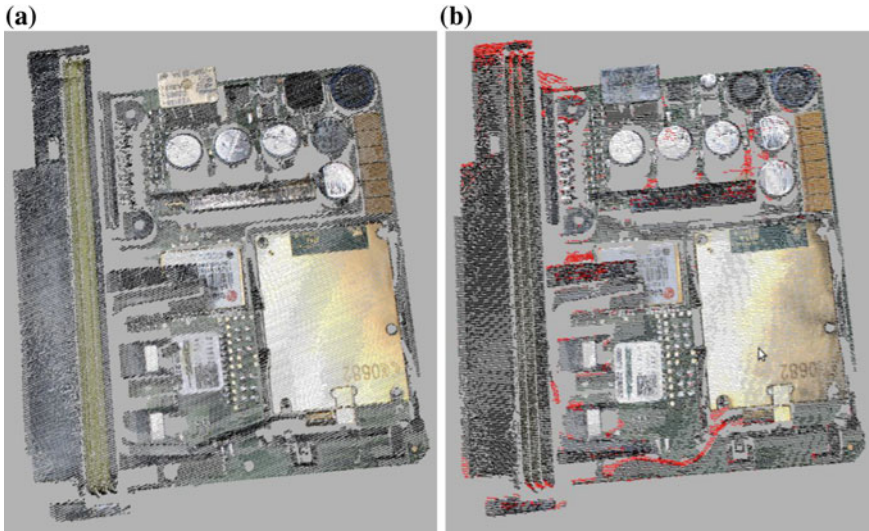
Fig. 7.24 Automatic change detection. **a** Reference point cloud. **b** Test point cloud with differences in *red*

an example of a point cloud acquired from a correctly assembled board that serves as the reference. Figure 7.24b shows the point cloud belonging to a board to be tested. The input point cloud can be carefully aligned with respect to the reference point cloud, so that its points that do not have a correspondence in the reference model can be identified. A point is considered to have a correspondence in the reference cloud if its distance to the reference cloud is under a user-defined threshold. The use of a color point cloud also allows the distances to be computed in the XYZ-RGB or XYZ-HSV space, instead of the XYZ space. By using an octree data structure to represent the point cloud, a very fast computation of point neighbors is obtained. As it can be noticed in Fig. 7.24b, where the points without correspondence are colored in red, the black cable at the bottom of the figure is detected as misplaced. Some small differences between the two point clouds due to outliers or border effects are also detected, but these can be easily removed by filtering out small differences.

7.7 Industrial Application

The hardware and software presented in this work also led to a new industrial solution called *Smart Check 3D*, produced and sold by IT+Robotics srl. It allows 3D completeness inspection of products on a production line. Smart Check 3D is available either as flexible scanning system mounted on a manipulator arm (Fig. 7.25a) or as inspection tunnel within a production line which allows to reach the maximum scanning speed.
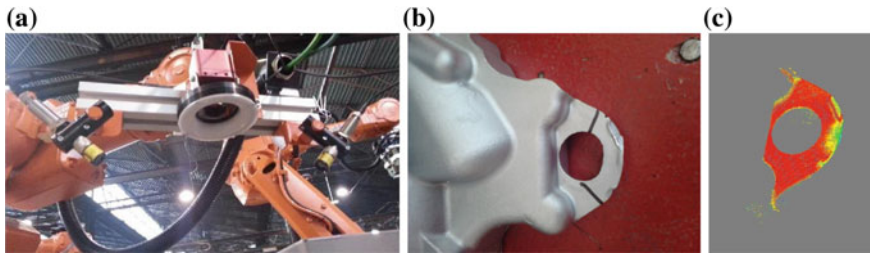
Fig. 7.25 **a** Industrial version of the dual-laser triangulation system proposed in this work. **b** Part to be scanned. **c** 3D point cloud resulting from the scan with color-coded quality inspection

In Fig. 7.25b, a picture of an object which had to be inspected by our system is reported, while the 3D point cloud obtained from a single scan is reported in Fig. 7.25c. Red color highlights the part of the scan which is considered as consistent with the CAD model of the object, while yellow and green colors show the parts where there are deviations from the model, thus those parts which are defective.

7.8 Conclusions

In this chapter, we presented the 2.5D color scanner we developed for real-time 3D color reconstruction of objects within a production line for the purpose of completeness inspection in 3D. We also detailed the calibration procedure that allowed us to reach very high accuracy in the reconstruction without the need for complex calibration objects. The wide range of output products that our software can provide proved to be very effective for performing inspection tasks which would be extremely difficult to solve for state-of-the-art 2D-based methods. Finally, we show that our system has already been integrated into a real industrial solution.

As a future work, we envision to also test our system with cameras with onboard processing, which could further improve the scanning speed.

References

1. Bi Z, Wang L (2010) Advances in 3d data acquisition and processing for industrial applications. Robot Comput Integr Manufact 26:403–413
2. Chen C, Kak A (1987) Modeling and calibration of a structured light scanner for 3-d robot vision. Int Conf Robot Autom 4:807–815

3. Huynh DQ, Owens RA, Hartmann PE (1999) Calibrating a structured light stripe system: a novel approach. Int J Comput Vis 33:73–86
4. Malamas EN, Petrakis EGM, Zervakis M, Petit L, Legat J-D (2003) A survey on industrial vision systems, applications and tools. Image Vis Comput 21:171–188
5. McIvor AM (2002) Non-linear calibration of a laser stripe profiler. Opt Eng 41:205
6. Pears N, Liu Y, Bunting P (eds) (2012) 3D imaging, analysis and applications. Springer, London
7. Reid ID (1996) Projective calibration of a laser-stripe range finder. Image Vis Comput 14:659–666
8. So EWY, Munaro M, Michieletto S, Menegatti E, Tonello S (2013) 3d complete: efficient completeness inspection using a 2.5d color scanner. Comput Ind Spec Issue 3D Imaging Ind Elsevier 64(9):1237–1252
9. So EWY, Michieletto S, Menegatti E (2012) Calibration of a dual-laser triangulation system for assembly line completeness inspection. In: IEEE international symposium on robotic and sensors environments, pp 138–143
10. Stocher W, Biegelbauer G (2005) Automated simultaneous calibration of a multi-view laser stripe profiler. In: IEEE international conference on robotics and automation, pp 4424–4429
11. Trucco E, Fisher RB, Fitzgibbon AW, Naidu DK (1998) Calibration, data consistency and model acquisition with laser stripers. Int J Comput Integr Manuf 11:292–310
12. Vilaça JL, Fonseca JC, Pinho AM (2009) Calibration procedure for 3d measurement systems using two cameras and a laser line. Opt Laser Technol 41:112–119
13. Yamauchi K, Saito H, Sato Y (2008) Calibration of a structured light system by observing planar object from unknown viewpoints. In: International conference on pattern recognition, pp 1–4
14. Zhou F, Zhang G (2005) Complete calibration of a structured light stripe vision sensor through planar target of unknown orientations. Image Vis Comput 23:59–67

Chapter 8
X-ray Computed Tomography for Non-destructive Testing and Materials Characterization

Johann Kastner and Christoph Heinzl

Abstract X-ray computed tomography (XCT) with flat-panel matrix detectors is a promising non-destructive method to scan and characterize external and internal structures of a specimen. Using XCT, both size and topology of different types of heterogeneities may vary largely. The different size scales of the heterogeneities and of the affected material volume require appropriate tomographic methods, scanning geometries and resolutions. In order to quantify features of interest from XCT scans, a major challenge is still found in the analysis and visualization of the generated XCT data: Advanced 3D-image processing techniques are needed for robust extraction and characterization of each single feature of interest. Interactive visualization techniques are needed for further exploration and analysis of these increasingly complex data spaces in order to provide novel insights for domain specialists. As examples of how image processing and visualization complement XCT various application cases are presented.

Contents

J. Kastner (✉) · C. Heinzl
University of Applied Sciences—Upper Austria, Stelzhamerstrasse 23, 4600 Wels, Austria
e-mail: Johann.Kastner@fh-wels.at

8.1 Introduction and Applications of X-ray Computed Tomography

Non-destructive testing (NDT), non-destructive evaluation (NDE) and non-destructive inspection (NDI) are terms describing methods for testing without causing damage. By definition, non-destructive techniques are means, which allow inspections of materials and structures without destroying them and without impairing their future usefulness. The objective of each NDT method is to provide information about the following main material or sample parameters [1–6]:

- Isolated discontinuities (such as cracks, voids, inclusions, delamination, porosity) and material structure
- Dimensional measurements (thickness, diameter, shape and connectivity of the detected inhomogeneity, three-dimensional geometry of internal and external geometry)
- To a certain extent also physical and mechanical properties (physical density, crack growth, wear) and chemical composition (alloy and phase identification, impurities)

Usually, 1D and 2D methods are used but in recent years a wide variety of new three-dimensional NDT methods providing 3D information about specimens have been developed [1–6]. The result of these 3D methods is a volumetric 3D representation of the specimen. Due to the rapid progress in computer science and data processing and also the tremendous improvements in hardware performance, the efforts necessary to apply tomographic methods have decreased.

Laboratory cone beam X-ray computed tomography (XCT) is the most promising method for non-destructive and fully three-dimensional characterizations of materials and components [7–13]. The principle equipment of cone beam XCT is depicted in Fig. 8.1 and consists of an X-ray source, a turntable and a matrix detector. At each angular position of a 360° turn, a 2D projection image of the specimen is obtained, which represents the X-ray attenuation of the specimen in the form of grey values. The complete series of projection images allows a full
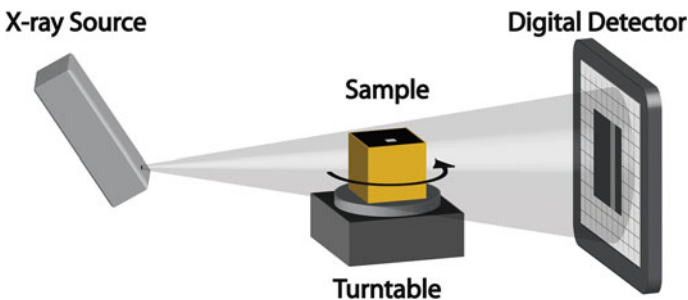
Fig. 8.1 Main equipment of cone beam X-ray computed tomography: X-ray source (*left*), rotary plate with specimen (*centre*), matrix detector (*right*)

reconstruction of the three-dimensional distribution of the spatial X-ray attenuation, resulting in a volumetric grey value dataset of the measurement area. In recent years, industrial XCT devices have been continuously improved to achieve higher resolutions, higher X-ray energies and higher accuracies.

The invention of XCT is one of the greatest innovations in the field of medical diagnostics in the last 100 years. The initial use of computed tomography during the 1970s sparked a revolution in the field of medical science because of the enormous benefit of non-invasive diagnostics. Medical X-ray tomography is well described in the literature, e.g. see Hsieh [14] and Buzug [15] and still a highly active topic of research.

In contrast to the field of medicine, XCT for industrial applications was not established as quickly. One of the first users in Europe outside the field of medical diagnostics was the Messerschmidt-Boelkow-Blohm GmbH (MBB, now Airbus helicopters) company in Munich. MBB has successfully applied conventional medical XCT scanners for NDT of industrial components such as rotor blades for helicopters [16] since 1979. The Federal Institute for Materials Research and Testing (BAM) in Berlin started the development of an XCT system for industrial applications back in 1978. The start-up of a corresponding system, equipped with a radionuclide source, was in 1983. Among other applications, this device was used to investigate barrels of nuclear waste [17].

But up to the 1990s, industrial XCT was either used in special cases only when conventional testing methods failed, or it was considered as a complementary method in the field of NDT. The breakthrough for industrial XCT in Europe took place much later at the end of the 1990s when the potential of XCT for three-dimensional geometrical measurements was recognized. An important pioneer in this application was the European automotive industry, which applied XCT for the first part testing of engine components made of aluminium castings. The following reasons are mainly accountable for the success of industrial XCT exploiting new application areas: The rapid development of high-performance computers and software for processing and visualization of 3D datasets allowed for analysing the steadily growing datasets. Furthermore, new solid-state flat-panel matrix detectors with a high dynamic range on the basis of amorphous silicon crystals increasingly replaced X-ray image amplifiers.

XCT has gained increasing importance and acceptance for the examination and inspection of industrial specimens in recent years. XCT finds its use in many different industrial fields, such as:

- Material science,
- Aerospace,
- Automotive,
- Electronics,
- Medical/food technology,
- Archaeology,
- Security, and
- Science and Research.
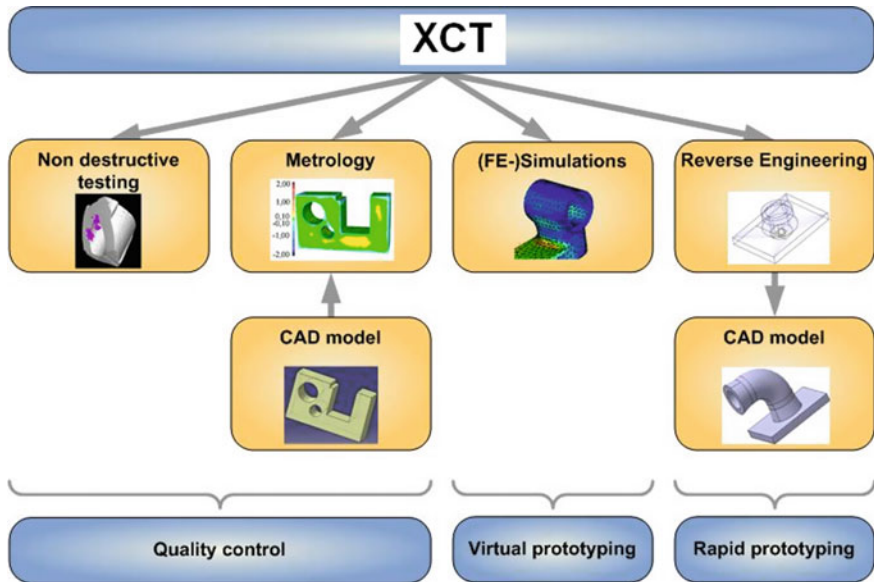
Fig. 8.2 Applications of cone beam XCT in the areas of quality control (non-destructive testing/materials characterization, metrology/dimensional measurement), virtual and rapid prototyping

The main application areas of cone beam XCT are shown in Fig. 8.2. Since XCT is applicable for a widespread variety of materials and the achievable resolutions are adequate for most applications, XCT is the method of choice for many applications:

- The original and still the most important application area of XCT is NDT and materials characterization. While for NDT internal structures such as shrink holes, material inclusions or cracks, are of special interest, in materials science the 3D characterization of different phases is relevant.
- The second major application area of XCT is metrology for dimensional measurements of 3D geometry features. The primary interest in metrology is the measurement of crucial distances, wall thicknesses, diameters and the 3D geometry of inner and outer surfaces in general. XCT therefore plays a large role in quality control.
- XCT is also beneficial in the field of virtual prototyping. In virtual prototyping using XCT, digital models of scanned prototypes are generated. XCT for virtual prototyping enables improvements of the geometry of prototypes in the pre-production phase. In addition, modelling and calculation of material properties can be performed based on XCT data, e.g. for the design of filter media or for the design of fibre-reinforced polymeric parts.
- Another important application area of XCT is reverse engineering and rapid prototyping: A computer aided design (CAD) model is extracted from an XCT

scan, which allows for reproductions and duplications of components by means of rapid prototyping methods.

- Further applications of XCT are for example found in the fields of in situ investigations, in biomedical sciences, geology, arts or archaeology.

Besides NDT/materials characterization, the second major application of XCT is dimensional measurement (metrology), which is a relatively new field of research [18]. The primary interest in metrology is the measurement of inspection features such as crucial distances, wall thicknesses and diameters. In quality assurance, metrology is a very common method which measures the surface geometry of a component for geometric tolerance using tactile or, more recently, optical sensors. XCT is an alternative to coordinate measuring machines and in addition it obtains not only surface information but also accurate information about internal structures which are not accessible to tactile measuring systems. Furthermore, XCT even allows characterizing partly or fully assembled components.

A common technique for the visualization of geometrical deviations using colour coding is actual/nominal comparison: A comparison for an Al-die casting part is shown in Fig. 8.3 (left picture). The colour scale denominates the deviations between the surface extracted from XCT data (isosurface) and the reference data (i.e. the CAD model). Each position of the reference object is coded with a colour corresponding to the local deviation.

In quality assurance, metrology is very important, because the accurate and reproducible measurement of the component's geometry as basis for production is a highly relevant task for many industrial branches. Inspection of geometric tolerances can be a complex task, because geometric errors are related to three-dimensional features and estimating this type of error is usually based on a cloud of points that has
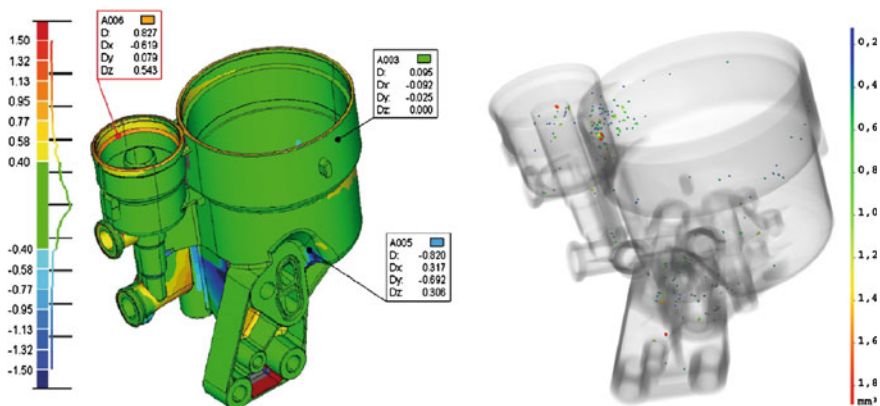
Fig. 8.3 Main application areas of XCT demonstrated with the help of an Al-die-casting part: (*left picture*) Metrology: Actual/nominal comparison between CT data and CAD model (*colours* correspond to the local deviation in mm); (*right picture*) non-destructive testing: semi-transparent 3D porosity analysis (*pores* are coloured *blue*, *green*, *yellow* and *red*)

to be measured on the machined surface. Coordinate measurement machines (CMMs) are the most common equipment for 3D measurement, because of their accuracy and flexibility.

XCT is an alternative to coordinate measurement machines, because XCT obtains not only surface information but also accurate information about inner and outer geometries which are not accessible to tactile or optical measuring systems. XCT is an especially promising technique for metrology purposes where a lot of measurement features have to be inspected, e.g. in first part inspection. By XCT, the inspection of all quality-relevant features such as diameter, distance, angle, roundness, and evenness can be performed within the generated 3D model. Here, one property of XCT is very important: in contrast to CMM, the measurement time of XCT does not depend on the number of features to be measured [14], which is a big advantage. For XCT typical overall analysis times for injection-moulded or light metal parts including the set-up of the work piece, the scanning and the feature extraction are around 60 min. XCT can reduce the inspection time considerably. The applications of XCT for metrology are in the following different fields:

- 3D geometry measurement
- Measurement of hidden structures and inner geometries;
- Measurement of wall thicknesses;
- Actual/nominal comparison which is especially important for first article inspection;
- Generation of STL representations (stereo lithography) of the surface of the specimen;
- Generation of CAD files of the component (reverse engineering); and
- XCT as basis for duplication by rapid prototyping methods.

Currently, XCT is not yet commonly established as a standard measurement method in metrology. One requirement for this to happen is the knowledge of the underlying measurement uncertainty, which is indispensable for comparability and acceptance of measurement results. In XCT, the measurement process is influenced by the X-ray source, the specimen, the detector, the scanning geometry, the environment, the user, the reconstruction and the data evaluation, as well as many further aspects. Every parameter of the whole process chain as listed below has a significant influence and needs to be considered appropriately and accurately:

- Calibration/adjustment of the XCT device: for this task, various reference standards such as a ball bar, a ball plate featuring spherical calottes, a calotte cube or other parts can be used;
- Scanning parameters: more than 40 quantities are relevant;
- Reconstruction: the reconstruction algorithm itself and the used reconstruction parameter are relevant;
- Reduction or correction of artefacts;
- Large data processing, in particular preprocessing of XCT data and proper segmentation/surface extraction; and
- Extraction of measurement features from the processed XCT data.

Despite of the widespread applications of XCT, for all applications artefacts may be a critical issue affecting the quality of the resulting dataset as well as subsequent data evaluations [19]. Artefacts are defined as artificial structures in the XCT data which do not correspond to real structures in the specimen. Artefacts may be introduced for various reasons, e.g. regarding scanning parameters, specimen geometry, reconstruction parameters or environmental conditions. In particular industrial 3D XCT systems based on cone beam geometry and matrix detectors are prone to artefacts such as noise-induced streaks, aliasing, beam hardening, partial volume and scattered radiation effects [14].

8.2   X-ray Computed Tomography Technology

8.2.1   Generation of X-rays and Interaction with Matter

The conventional method of producing X-rays in a laboratory is to use an evacuated tube with an acceleration voltage of up to 800 kV [20]. The tube contains a cathode filament where an electron cloud is formed and these electrons are subsequently accelerated towards the anode, which integrates the target material for the generation of X-rays. The anode is typically made of tungsten or molybdenum due to their convenient X-ray energy spectrum. In order to allow for high thermal loads, the targets typically get cooled (e.g. using water-cooling), are rotated (e.g. rotating anode) or embedded within materials with fast heat transport (e.g. tungsten on diamond substrate). When the accelerated electrons interact with the target material, different types of collisions take place and generate an X-ray spectrum with a behaviour shown in Fig. 8.4. The number of photons per second and per 100 μA target current related to the solid angle (steradiant) is plotted against the energy.
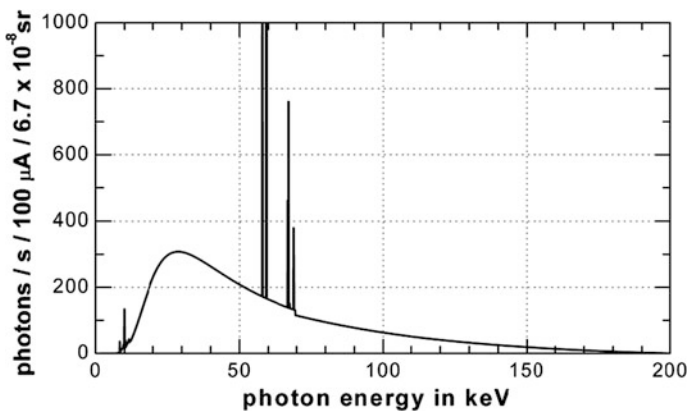
Fig. 8.4 Simulated X-ray energy spectrum of an X-ray tube with a tungsten target operating at 200 kV [21] (Energy channel bandwidth ΔE is 100 eV). The spikes correspond to the characteristic radiation of the tungsten target [22]
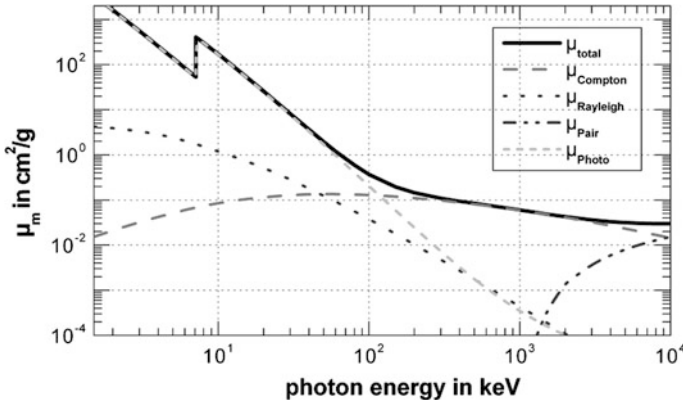
Fig. 8.5 Total mass attenuation coefficient (μ_{total}) for iron against the photon energy decomposed into the Rayleigh scattering ($\mu_{Rayleigh}$), Compton scattering ($\mu_{Compton}$), photoelectric absorption (μ_{Photo}) and pair production (μ_{Pair}). For X-ray tubes, relevant areas are up to 600 keV. The data were taken from NIST Standard Reference Database [25]

The X-rays are known to have a very high, material-dependent capability of matter penetration. Several fundamental ways in which X-ray photons interact with matter are known: The most important physical mechanisms are photoelectric absorption, Rayleigh or Thomson scattering, Compton scattering and pair production which are explained by Moore [23] and Eisberg [24].

Most of the mechanisms of absorption involve the conversion of the photon's energy to another form, while some simply change the photon's direction. For each mechanism, the probability that photons interact with the material can be defined in a single attenuation coefficient. The attenuation coefficient is a function of the photon energy (E), the mass density (ρ) and the atomic number (Z) of the transmitted material. The total probability of interaction is the sum of the probabilities of each physical mechanism, which is shown as example in the bold line in Fig. 8.5 for iron. The individual interaction probabilities are strongly dependent on the energy of the incident photon. While for small energies below 100 keV, photoelectric absorption is the dominant effect, for energies above 100 keV the total attenuation is almost entirely caused by Compton scattering interaction. Regarding the probability of interactions, Rayleigh scattering is relevant but not dominant. Pair production, however, has a relevant contribution only for energies above 1 MeV and can therefore be neglected for X-ray tubes.

The detection of X-ray photons is based on the same principles of interactions between X-ray radiation and material. The purpose of a detector is to convert a photon flux into an electrical signal. Apart from gas detectors, there are numerous different types of digital X-ray detectors, such as image intensifiers, flat screen detectors, linear detector arrays and X-ray sensitive CCD cameras. For XCT, the most important and common ones are digital flat-panel detectors. These detectors can be divided into two types: direct and indirect conversion detectors. Both types

are based on thin layers of amorphous silicon (a-Si) deposited onto glass substrates, with two-dimensional arrays of detector elements fabricated on the a-Si. For the indirect detection systems, X-ray photons pass into a phosphorescent screen (usually Gd_2O_2S:Tb or CsI:Tl) and produce visible light, which is in turn detected by the array of thin film transistor (TFT)-switched photodiodes. For detectors with direct conversion, the radiation is directly converted in the sensor layer, which is based on Se, CdTe or ZnCdTe [12, 26, 27]. A novel and promising type of detectors aims to generate noise-free data by photon counting: Medipix-type detectors target for high spatial resolution and high contrast and provide single photon counting capability. Medipix-based detectors can be combined with different semiconductor sensors, converting the X-rays directly into detectable electric signals.

8.2.2   Principle of Cone Beam XCT

Due to the measurement speed and quality, XCT systems with cone beam geometry and flat-panel matrix detectors [28, 29] have gained general acceptance in the market of XCT systems for industrial applications and material characterization. Using XCT, a specimen is placed on a high precision rotating stage between the X-ray source and a flat-panel detector. The X-ray tube produces a conic beam of photons that penetrate the specimen to be analysed. Due to this geometry, a magnified projection image (2D radiograph) is recorded on the digital detector. Reducing the source-to-object distance by moving the rotating stage with the specimen allows for a larger magnification and an improved spatial resolution, but it also reduces the field of view. The specimen is rotated step by step, taking projection images at each angular position. One scan usually covers a step width of fractions of a degree (up to one or more degrees for fast scans) and a full rotation of 360°. Furthermore, for specific applications also limited angle scans [30, 31] may be performed. Feldkamp et al. [32] describe an approximating reconstruction algorithm to retrieve 3D data from a series of 2D radiographs for circular cone beam tomography. It is an approximation in the sense that the reconstruction result will deviate somewhat from the measured object regardless of the measurement resolution. For moderate cone angles, these differences are however small and often acceptable [33, 34]. The simplicity of this method has made it the most used algorithm for cone beam reconstruction. The result of the reconstruction algorithm is a volumetric dataset, which consists of discrete volume elements. A volume element is called a voxel and represents the fusion of *volume* and *pixel*. Each voxel represents a grey value corresponding to the effective X-ray attenuation coefficient at each position of the dataset.

The Upper Austrian University of Applied Sciences operates two cone beam XCT devices: a dual-source XCT device and a sub-micro XCT device (Fig. 8.6). The XCT system Rayscan 250XE is equipped with two X-ray sources for the inspection of a wide variety of objects, ranging from micro-parts to cylinder heads. This system is equipped with a 225 kV micro-focus and a 450 kV normal-focus tube to add flexibility to the system. The target material is tungsten for both tubes.
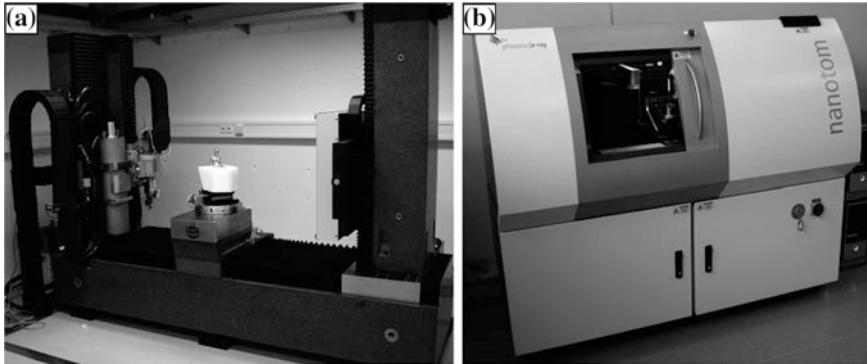
Fig. 8.6 XCT devices available at the Upper Austrian University of Applied Sciences: **a** Rayscan 250XE with a 225 kV micro-focus and a 450 kV normal-focus tube. **b** Desktop device GE nanotom 180NF with a 180 kV sub-micro-focus tube encased in a radiation safe cabinet

The detector is a 16-bit solid-state 2048×2048 pixels a-Si flat-panel detector. The source-to-detector distance is 1.5 m, the maximum sample diameter is 0.6 m and maximum sample height is 2 mm. The typical duration of XCT scans with this device is 30 min–1 h. Figure 8.6b shows the desktop XCT device GE nanotom 180NF which uses a 180-kV sub-micro-focus tube with an exchangeable trans-mission target and a flat-panel detector with 2316×2316 pixels. The sub-micro-focus tube allows a maximum voltage of 180 kV with a maximum output of 15 W. The X-ray spot size varies from a few microns to 0.8 µm. The GE nanotom 180NF is a fully protected device and features a measurement chamber with inherent protection, which shields X-ray radiation by lead walls, so that no further means of protection is necessary. The source-to-detector distance is about 0.5 m, and samples with a maximum diameter of 120 mm can be scanned. Representative scan times are 1–4 h.

8.2.3 Multi-Scale Tomography

The different size scales of the heterogeneities and of the affected material volume require appropriate tomographic methods (see Fig. 8.7): macro-XCT or standard-XCT, (>10 µm); micro-focus X-ray tomographs µ-XCT, (currently >3 µm); sub-µm XCT, (currently >0.4 µm); and synchrotron tomography s-XCT, (currently >0.2 µm) [35]. s-XCT applying KB-Mirrors have reached resolutions down to $(0.04 \ \mu m)^3$ at ESRF/ID2237. Focussed ion beam tomography (FIBT) represents a destructive method and can be used to identify heterogeneities in a size range of 10 nm–1 µm18. The indicated resolution limits vary with the maximum transmitted distance, which limits the specimen size to about $1000 \times$ resolution limit.

Figure 8.8 shows dimensional scales for voxel size and object size (diameter) for the considered tomographic methods. Macro-, µ- and sub-µm XCT (also called
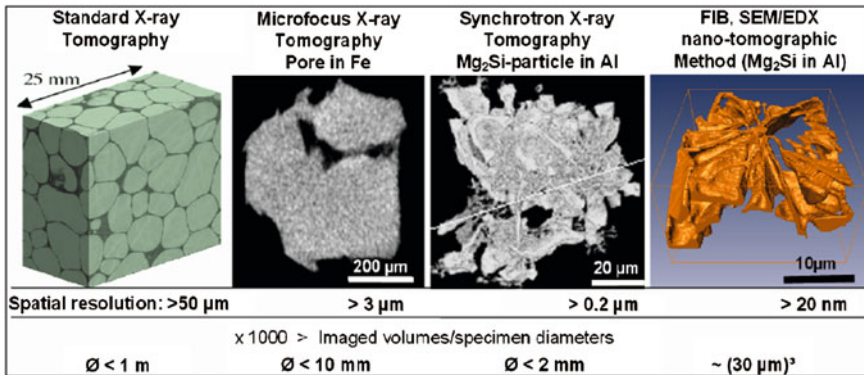
Fig. 8.7 Multi-scale X-ray tomography showing the different resolutions and object sizes accessible by cone beam XCT, by s-XCT and FIB tomography. *FIB*—focussed ion beam, *SEM*—scanning electron microscopy, *EDX*—energy dispersive X-ray analysis [36]

Fig. 8.8 Multi-scale X-ray tomography showing the different resolutions and object sizes accessible by cone beam XCT systems of FHW (macro-XCT, μ-XCT, nano-XCT/sub-μm XCT), by s-XCT and by FIB tomography
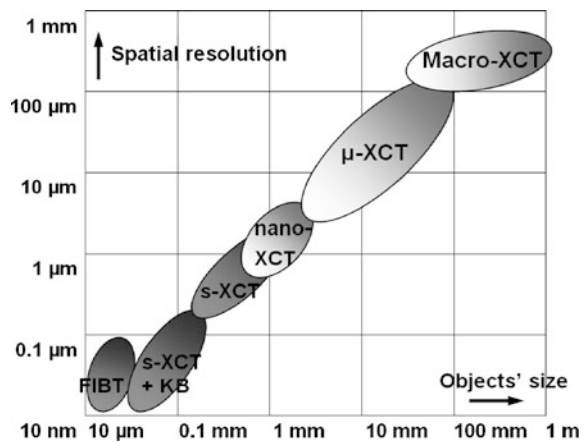
nano-XCT) systems are available at the computed tomography research group at the University of Applied Sciences—Upper Austria. They cover a range of more than four orders of magnitude for both parameters.

8.3   Image Processing and Visualization of XCT Data

8.3.1   Preprocessing and Segmentation

As an essential prerequisite of feature extraction, accurate segmentations and evaluations of the objects of interest in the XCT dataset are necessary. Due to the fact that XCT scans are affected to a certain degree by noise and artefacts, image
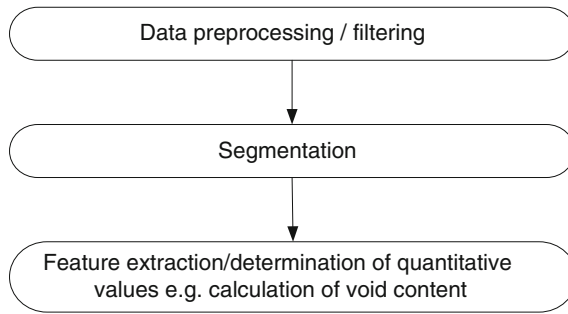
Fig. 8.9 Typical XCT data evaluation workflow consisting of several steps. The input dataset is the reconstructed XCT dataset. The output dataset contains quantitative values such as void content, porosity or a surface model of the investigated component

enhancement by applying (a series of) adequate filters is often necessary. A typical XCT evaluation workflow consists of several steps. A simple XCT data evaluation workflow with three steps is shown in Fig. 8.9.

After reconstruction, the first step of preprocessing is frequently found in reducing artefacts and noise to increase contrast between different features. For these tasks, different artefact correction methods (e.g. IAR) can be used and a prefiltering/denoising step, e.g. Gaussian filtering or anisotropic diffusion filters can be applied. The term 'denoising' describes the attempt to remove noise and retain the signal regardless of the frequency content of the signal. XCT images can be enhanced by using multi-resolution image denoising methods. Recent studies benchmark denoising algorithms on XCT data [37]. A precondition for the correct application of denoising filter methods is an exact noise characterization of the imaging system (e.g. Poisson noise, Gaussian noise). However, if the photon transfer curve of the XCT detector is measured or known, multi-resolution denoising methods (in particular wavelet filters and the platelet filter) can improve the quality of XCT images considerably. This is demonstrated in Fig. 8.10.

For segmentation, different established and novel methods are used. The simplest grey value-based segmentation method is thresholding using either global or local image information. A simple method to binarize an image is to select a grey value with the visual support of a grey value histogram and a preview showing the binarized result on one slice. Apart from visual determination of the appropriate threshold, automatic histogram-based thresholding methods (e.g. Otsu [38]) are available. All these techniques suffer from the fact that they rely on grey value information. To analyse local features, region-based methods are more suitable. For example, region growing starts from seed points (usually provided by the operator) and neighbouring voxels are examined one at a time and added to the growing region if they are sufficiently similar. A comparison may be made with the entire region or just with the local pixels (voxels), with the latter method allowing gradual variations in brightness. The procedure continues until no more pixels (voxels) can
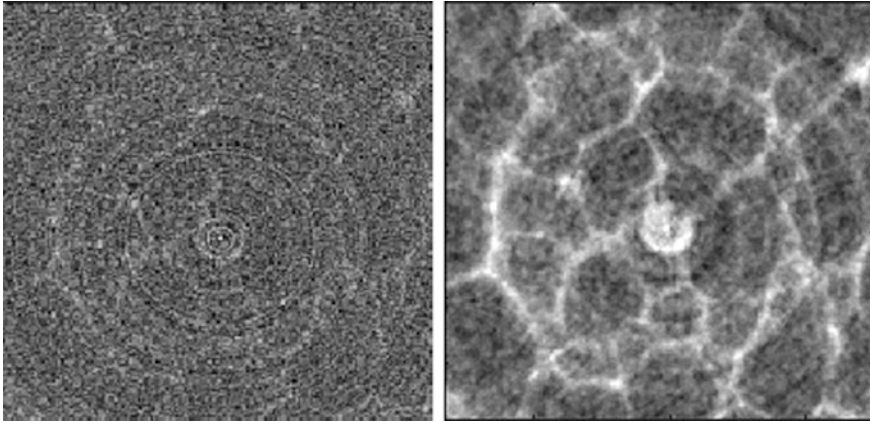
Fig. 8.10 Result of denoising process (polymeric foam structure), *left* original XCT picture; *right* platelet denoised. The concentric rings are the so-called ring artefacts

be added. Despite their simplicity, simple methods as thresholding or region growing can be very useful for initialization of a more sophisticated segmentation pipeline, e.g. using watersheds or levelsets.

The following sections describe image processing and visualization of XCT data on two application cases: the characterization of metals and the characterization of polymeric materials.

8.3.2   Characterization of Metals

There are many publications that deal with XCT of light metals and other metals. As an example, this section shows the possibilities of XCT applied on high-strength AlZnMgCu alloys [39]. High-strength wrought aluminium alloys are widely used in aerospace vehicles because of their high specific stiffness and strength, good formability and resistance to corrosion. Chill cast Al-Zn-Mg-Cu–billets contain Sc and Zr, which form secondary $Al_3(Sc,Zr)$ dispersoids during homogenization, but as well primary $Al_3(Sc,Zr)$ in the melt. In addition to these particles, interdentritic segregations rich in highly absorbing elements such as Zn and Cu appear (inter-dendritic eutectic containing Zn_2Mg and Cu-rich phases). Figure 8.11 shows the workflow for differentiation between interdentritic Cu- and Zn-rich regions and $Al_3(Sc,Zr)$ particles in the XCT data of $AlZn_6Mg_2Cu_2$. It starts with correction by iterative artefact correction. The second step is segmentation by a global threshold method and the last step is differentiation between interdentritic Cu- and Zn-rich regions and $Al_3(Sc,Zr)$ particles by using the form factor F. The form or shape factor F represents three-dimensional sphericity. It is calculated by (Eq. 8.1) of each individual heterogeneity
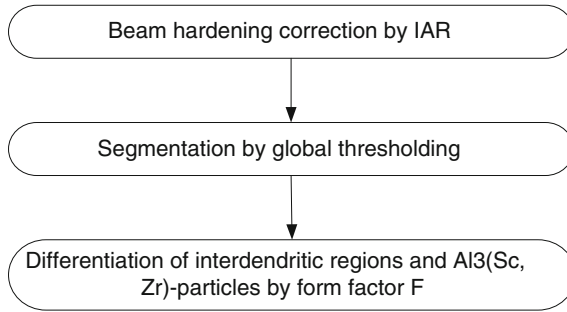
Fig. 8.11 Workflow for differentiation between interdentric Cu- and Zn-rich regions and Al$_3$(Sc, Zr) particles in the XCT data of AlZn6Mg2Cu2. *IAR* iterative artefact reduction

$$F = 6 \cdot \sqrt{\pi} \cdot \frac{\text{volume}}{\sqrt{\text{surface}^3}}, \quad \text{where } F \in [0, 1], \text{ ideal sphere } F = 1. \quad (8.1)$$

A suitable threshold of $F = 0.4$ was applied to differentiate between the inter-dentritic regions and the Al$_3$(Sc,Zr) particles. This is shown in Fig. 8.12 for a selected volume of $0.27 \times 0.29 \times 0.12$ µm^3, where the segmented interdendritic Cu- and Zn-rich regions are visualized in orange and the Al$_3$(Sc,Zr) particles are presented in red and marked with Z. Here, $F < 0.4$ was chosen for the interdendritic regions and $F > 0.4$ was chosen for the particles.

The next example is a special Mg-alloy. Figure 8.13 shows a cross-sectional XCT picture of as-cast (left pictures) and extruded (right pictures) MgY7Zn2.5 [40].
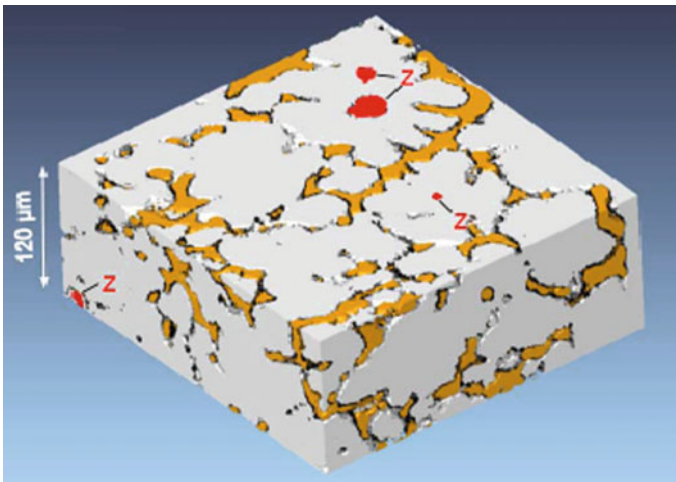
Fig. 8.12 Typical arrangements of α-Al dendrites (*white*), interdendritic eutectic regions (*orange*) and Al$_3$(Sc,Zr) particles (Z) in a volume of $0.27 \times 0.29 \times 0.12$ mm^3
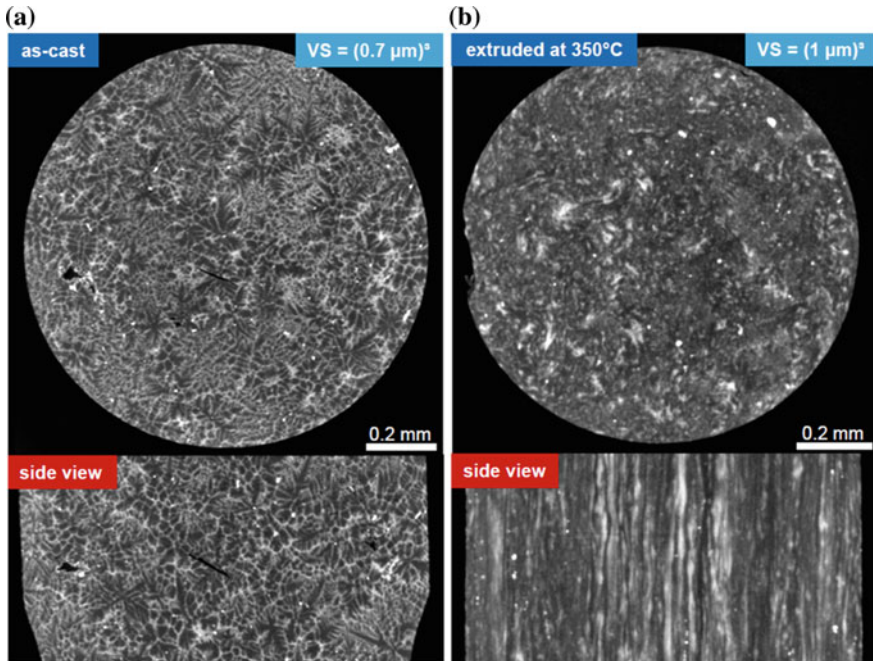
Fig. 8.13 XCT slices of MgY7Zn2.5: **a** as-cast and **b** extruded at 350 °C. The CT measurements were carried out at 60 kV with a voxel size of $(0.7\ \mu m)^3$ and $(1\ \mu m)^3$, respectively

Four individual phases can be discriminated: high-density particles (brightest particles), LPSO-phase (bright phase in the interdendritic region), Mg-matrix (dark) and features with grey values absorbing lower than those of the Mg-matrix, which can be identified as pores. The volume fraction of the individual phases can be discriminated by grey value segmentation. Thus, the volume fraction of high-density particles (hdP) in the as-cast condition is about 0.05–0.08 vol.%, the concentration of the LPSO-phase is around 25 vol.% and concentration of the pores is 0.01–0.3 vol.%.

MgY7Zn2.5 was subjected to tensile deformation at 350 °C up to different strain levels (0, 10, 30, 60, 100, 160 and 220 %). Figure 8.14 shows the phases found in three dimensions. By segmenting with a simple grey value-threshold-based method, the relative content and the interconnectivity of the phases can be evaluated. In Fig. 8.14, the Mg-matrix is illustrated transparently and the other phases green, red and blue. High-density particles are only situated within the LPSO-phase. The pores are oriented along the LPSO-phase in the tensile direction. The number and fraction of pores increases with increasing strain. A detailed quantitative evaluation of the CT results was carried out to determine the interconnectivity (interconnectivity of a phase, I = largest connected volume of a particle Vphase, largest divided by the whole volume of a certain phase Vphase, total) of the individual phases at elongations between 0 and 220 %.
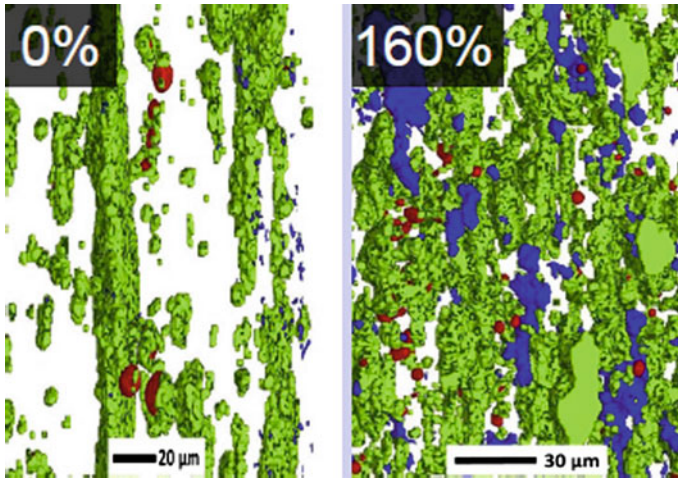
Fig. 8.14 Segmented and coloured XCT tomograms of MgY7Zn2.5: tensile deformation at 350 ° C up to different strain levels (0 and 160 %). The colour code is as follows: *transparent* Mg; *green* LPSO; *red* hdP—high-density particles; *blue* pores

Interconnectivity $_{Pores}$: 0.5–5.1
Interconnectivity $_{LPSO}$: 97.2–98.8

The volume fraction and the interconnectivity of the pores and LPSO-phase deepen on the elongation applied. It can be seen that the volume fraction and interconnectivity of the LPSO-phase does not depend on the elongation, whereas the pore fraction and the interconnectivity of pores increase with elongation.

8.3.3 Characterization of Polymeric Materials

Fibre-reinforced polymers (FRPs) are increasingly used in industry due to their superior characteristics over conventional materials (e.g. higher specific stiffness and strength, greater corrosion resistance, lower weight and shorter production times). in particular, properties such as fibre length and fibre orientation are important for the development of new FRP materials and components. Figure 8.15 shows two concepts to determine fibre orientation and fibre length distribution in a glass fibre-filled polymer. The first concept is based on segmentation using a global threshold. This binarization is followed by a detailed analysis of regions where fibres are touching. The second concept is based on analysis of the original grey value image. For each voxel, the local fibre orientation is determined by calculating the Hessian matrix and analysing their eigenvalues. The capability of the two data analysis concepts to extract orientation and length was investigated.

From the original XCT dataset, quantitative data can be extracted by applying the methods shown in Fig. 8.15. For instance, the XCT data of glass-fibre-reinforced
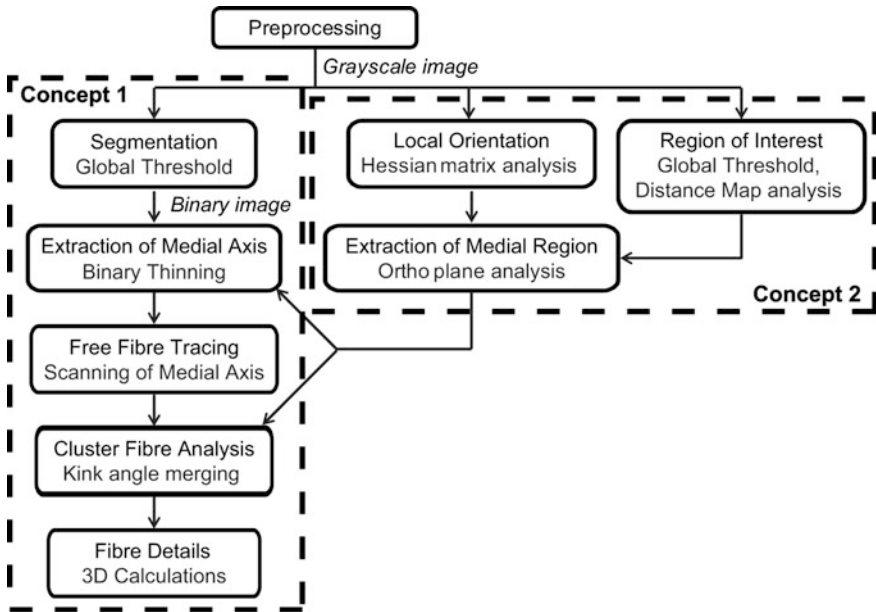
Fig. 8.15 Principle scheme of two workflows (concepts 1 and 2) for XCT data evaluation to determine fibre orientation and fibre length distribution in a glass-fibre-filled polymer [41]
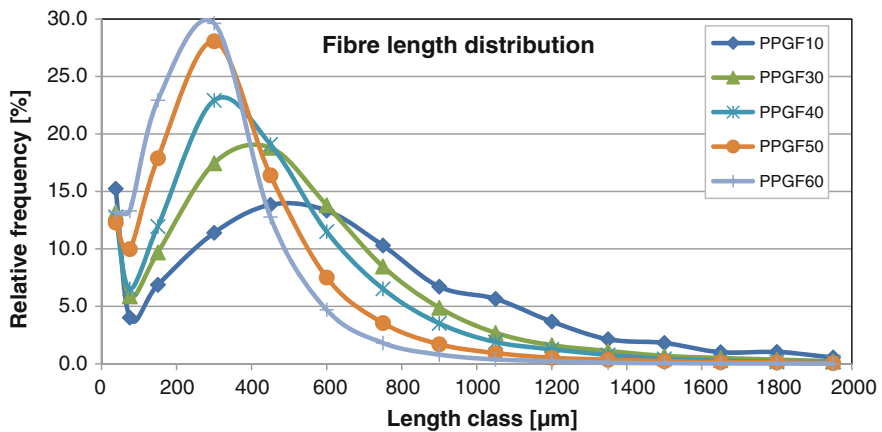
Fig. 8.16 Fibre length distribution function of injection-moulded polypropylene reinforced with 10, 30, 40, 50 and 60 % glass fibres. The fibre lengths were determined from XCT data

polymer can be processed to extract fibre orientation and fibre length distribution function. The XCT data are processed by several consecutive steps: preprocessing, segmentation, medial axis extraction, individual fibre extraction including cluster analysis and final fibre characterization. Figure 8.16 shows fibre length distribution
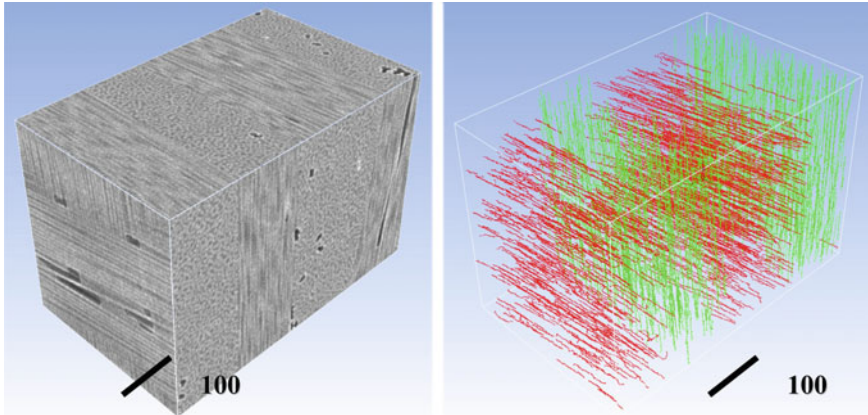
Fig. 8.17 XCT data of a carbon-fibre-reinforced polymeric volume of $(570 \times 356 \times 374)$ μm^3. The voxel size was $(1 \ \mu m)^3$. The orientation of the fibres is visualized by colour coding. The bars have a length of 100 μm

functions of injection-moulded polypropylene reinforced with 10, 30, 40, 50 and 60 %. It is clearly recognizable that the fibre length distribution function is shifted to lower values with increasing fibre content. The higher the fibre content, the more fibres will be broken during the die injection moulding process.

Similar results can be expected for carbon-fibre-reinforced polymers, even though the fibres are usually much thinner and the contrast is significantly lower. The left picture of Fig. 8.17 shows a 3D representation of the CT data of a carbon-fibre-reinforced polymeric sample (CFRP). The right picture shows the extracted fibres. The orientation of the fibres is visualized in green and red.

A major challenge to support material scientist in their daily work is to provide tools for analysis and visualization of the generated quantitative fibre data. In order to address this challenge, we have introduced tools allowing for interactive exploration and analysis of high-resolution XCT data of FRPs [42] (Fig. 8.18). More specifically, advanced visualization and analysis techniques such as parallel coordinates and scatter plot matrices are used to define and refine fibre classes of interest. This way a clusters of fibres with similar characteristics, e.g. length, orientation may be created and visualized. If parallel coordinates and scatter plot matrix are linked to 2D and 3D views, then those classes can be nicely visualized in the spatial context. Using a polar plot, an overview of the fibre orientation distribution in the data may be provided to the end-user. Using bar charts, fibre length distributions are visualized. To analyse classes of fibres on a more regional and more abstract level, the metadata and blob visualizations allow to encode fibre regions with similar characteristics.

Another important application in this area is the characterization of porosities in the fibre-reinforced polymers. In particular, the porosity of carbon-fibre-reinforced polymers is a very important topic for practical applications since there is a direct correlation between porosity and mechanical properties, such as shear strength.
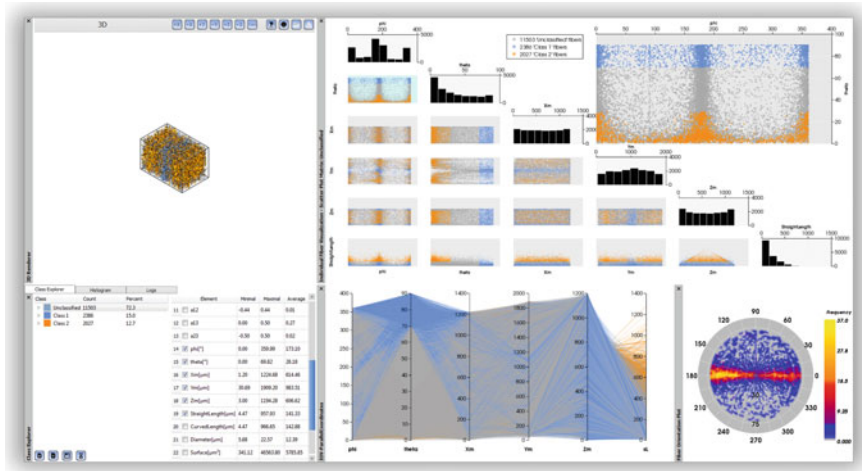
Fig. 8.18 Overview of the FiberScout system, an interactive tool for exploring and analysing fibre-reinforced polymers demonstrated on a glass fibre-reinforced polymer (GFRP) dataset [42]. The selected fibre classes are depicted in all views in *blue* and *orange* to explore and analyse their characteristics

Therefore, the porosity values of CFRP samples for practical use must be below a certain level, typically 2.5–5 % depending on the final application. The most common non-destructive method for determining porosity is ultrasonic testing, since there is a mostly linear correlation between ultrasonic attenuation and porosity. However, the ultrasonic attenuation coefficient depends not only on the porosity, but also on the shape and distribution of the pores and the presence of other material inhomogeneities, which can result in significant errors. Another very promising method for the non-destructive determination of porosity of composites is X-ray computed tomography [43]. Besides quantitative porosity determination, the visualization and the calculation of local pore properties, i.e. volume, surface, dimensions and shape factors, are also important for a NDT practitioner or a materials scientist. Since XCT gives a complete 3D representation of a CFRP specimen including inner and outer surfaces, CT data is a good basis for visualization of material properties and geometry. We have developed an approach for interactive exploration and visual analysis of CFRP specimens to enhance the evaluation of such components and materials. Our solution includes the calculation of local pore properties for pore classification and their visualization. Besides the porosity map view, the introduced pipeline (shown in Fig. 8.19) integrates region of interest and parallel coordinates interaction in a two-stage drill-down approach. The best viewpoint widget allows to see the quality of the viewpoints on a sphere. An example for the application of the interactive exploration and visual analysis of a certain volume of a CFRP sample is shown in Fig. 8.20. In the left image, a porosity
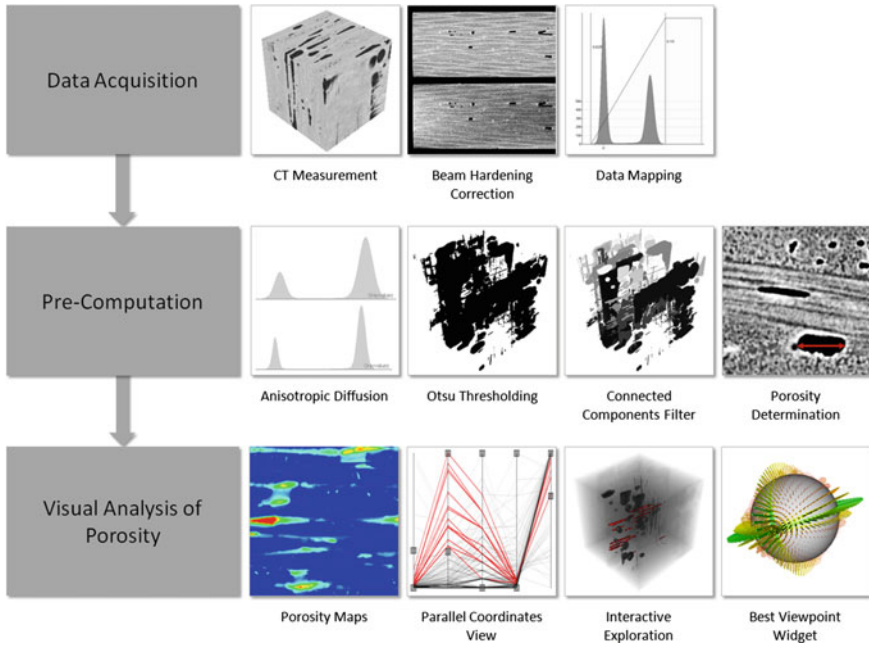
Fig. 8.19 Overview of the porosity visualization pipeline for interactive exploration and visual analysis of porosity in CFRP specimens [44]
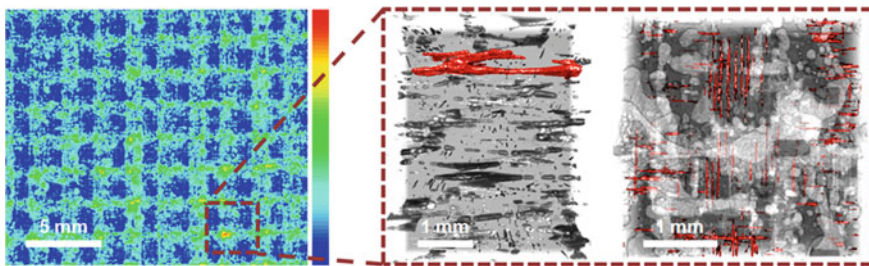
Fig. 8.20 Porosity map view of a carbon-fibre-reinforced polymeric specimen (*left*), where low porosity are coloured in *blue* and higher porosity in *green* and *red*. Porosity evaluation of a region of interest (*right*) showing different pore classifications using parallel coordinates interaction. The biggest pore in the region is selected and highlighted in the central picture. Only the small micropores between the fibres are highlighted on the right picture

map view of the specimen is depicted. The use of parallel coordinates allows to filter pores with specific pore properties. In the central picture of Fig. 8.20, the biggest pore in the selected region of interest is highlighted. The right picture of Fig. 8.20 shows long and thin micropores in the fibre bundles.
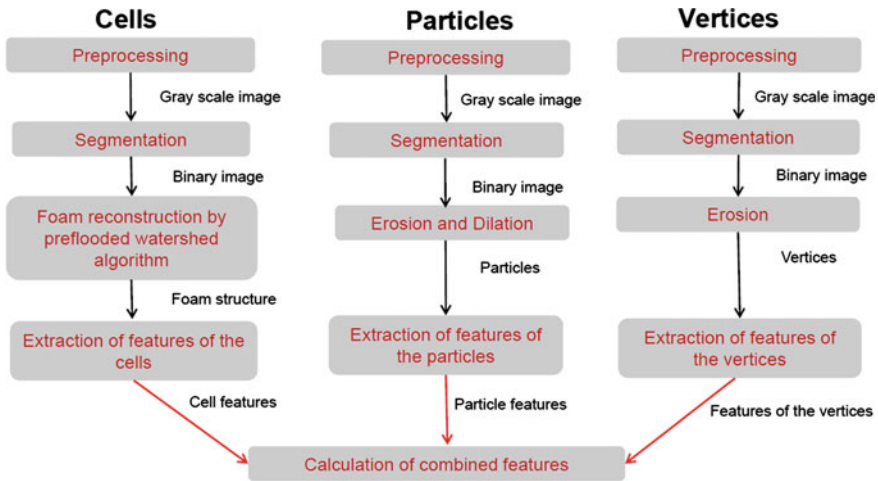
Fig. 8.21 Principle scheme of the workflows for XCT data evaluation of a cellulose-particle filled polymer foam to extract features for the cells, for the cellulose particles, for the vertices and combined features

For further filtering of pores in a specific direction, the dimensions bounds can be set in the parallel coordinates view additionally to the shape factor. Selected pores are highlighted in colour. Not selected pores are given in black where white and grey illustrate the typical twill weave pattern of the fibre bundles. Aside the properties of individual pores, also a cumulative representation of pore classes may be of interest for the end-user. To address this goal, mean objects may be calculated to visualize the mean pore object of user-defined classes [45].

Figure 8.21 shows a workflow for 3D microstructure characterization of a cellulose-particle-filled polymer foam. The XCT data were processed by various algorithmic steps (e.g. smoothing, thresholding, watershed transformation, erosion, dilation and feature extraction) to ascertain the three-dimensional open cell structure together with particle distribution, size and position [46]. Quantitative data for the cells (mean diameter, shape, volume and position), for the cellulose particles (mean diameter, shape, volume and position), for the polymer vertices (shape and position) and for combinations of them (e.g. distance between particles and vertices) were derived from the XCT data. Not only can fibre lengths be measured but also particle sizes can be extracted. This is shown in Fig. 8.22, where the particle diameter distribution function for cellulose particles of the polyurethane foam is presented. The limited CT resolution particles with a diameter below 10 µm were not considered. It can be seen that the maximum distribution function is at small diameters and the particle diameters reach values up to around 80 µm. The shape of the distribution function resembles an exponential function.
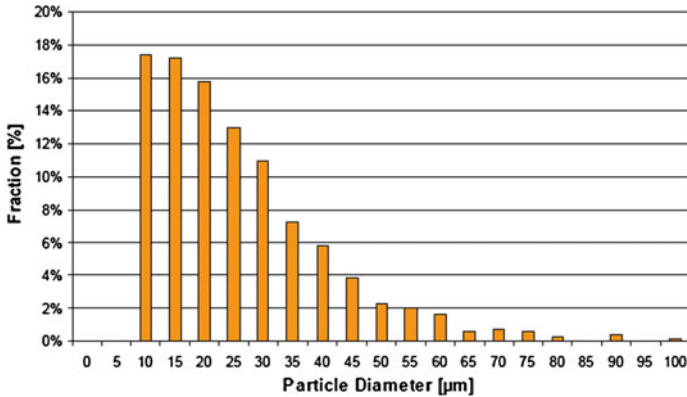
Fig. 8.22 Cellulose particle diameter distribution for the PU foam with cellulose particles

8.4 Conclusions

This chapter describes studies that deal with the detection and characterization of heterogeneities by means of XCT methods, namely cone beam XCT with micro-focus and sub-micro-focus X-ray sources. The main conclusions are as follows:

- X-ray computed tomography is a powerful NDT tool for characterization and visualization of different properties materials.
- Artefacts need to be carefully considered especially for quantitative data generated from XCT scans.
- Quantitative evaluation of material properties such as porosity determination, evaluation of fibre length and orientation distribution of individual fibres, and interconnectivity of phases are important applications of image processing methods applied to XCT data.
- The development of more advanced image processing and visualization methods is necessary to gain insight into the extracted quantitative data of XCT datasets for everyday industrial use.

References

1. Proceedings of 16th world conference on non-destructive testing, Montreal, 30 Aug–3 Sept 2004
2. Proceedings of 9th European conference on non-destructive testing, Berlin, 25–29 Sept 2006
3. Proceedings of 17th world conference on non-destructive testing, Shanghai, 25–28 Oct 2008

4. Proceedings of 10th European conference on non-destructive testing, Moscow, 7–11 June 2010
5. Proceedings of 18th world conference on non-destructive testing, Durban, 16–20 April 2012
6. Proceedings of 11th European conference on non-destructive testing, Prague, 6–10 Oct 2014
7. INSA-Lyon (ed) (2007) Proceedings of international symposium on digital industrial radiology and computerized tomography, Lyon, 25–27 June 2007
8. Kastner J (ed) (2008) Proceedings of Industrielle Computertomografietagung, Shaker Verlag, Wels, Austria, 26–27 Feb 2008 (ISBN 978-3-8322-6949-4)
9. Kastner J (ed) (2010) Proceedings of Industrielle Computertomografietagung, Shaker Verlag, Wels, Austria, 28–30 Sept 2010 (ISBN 978-3-8322-6949-4)
10. Kastner J (ed) (2012) Proceedings industrial computed tomography conference 2012, Shaker Verlag, Maastricht, 19–20 Sept 2012 (ISBN 978-3-8322-9418-2)
11. Kastner J (ed) (2014) Proceedings industrial computed tomography conference 2014, Shaker Verlag, Maastricht, 25–28 Feb 2014 (ISBN: 978-3-8440-2557-6)
12. Baruchel J, Buffiere JY, Maire E, Peix G (eds) (2000) X-ray tomography in materials science. Hermes Science Publications, Paris
13. De Chiffre L, Carmignato S, Kruth JP, Schmitt R, Weckenmann A (2014) Industrial applications of computed tomography. CIRP Ann Manufact Technol 63:655–677
14. Hsieh J (2003) Computed tomography, principles, design, artifacts and recent advances. SPIE The International Society for Optical Engineering, Bellingham
15. Buzug TM (2008) Computed tomography: from photon statistics to modern cone-beam CT. Springer, Berlin
16. Oster R (1999) Computed tomography as a non-destructive test method for fibre main rotor blades in development, series and maintenance. In: Proceedings of international symposium on computerized tomography for industrial applications and image processing in radiology (DGZfP-Berichtsband 67-CD)
17. Illerhaus B, Goebbels J, Kettschau J, Reimers P (1989) Non Destructive Waste Form and Package Characterization by Computed Tomography 12th Int. Symp. on the Scientific Basis for Nuclear Waste Management, Berlin, Oct. 10-13, 1988 Mat. Res. Soc. Symp. Proc. 127, 507-512
18. Heinzl Ch, Kastner J, Gröller E (2007) Surface extraction from multi-material components for metrology using dual energy CT. IEEE Trans Visual Comput Graphics 13(3):1520–1528
19. Heinzl C (2009) Analysis and Visualization of Industrial CT Data. Ph.D. thesis, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Favoritenstrasse 9-11/186, A-1040 Vienna, Austria
20. http://www.comet-xray.com/technology-en/high-energy. Accessed 15 Aug 2015
21. Ebel H (1999) Advances in X-ray Analysis. In: XRS 281999, No 4, pp 255–266
22. Harrer B, Kastner J (2011) Cone beam ct for non-destructive testing fabrication and characterization in the micro-nano range: new trends for two and three dimensional structures, 'X-ray microtomography: characterisation of structures and defect analysis'. Springer, Heidelberg, 119–150
23. Moore PO (ed) (2002) Nondestructive testing handbook, radiographic testing, vol 4. American Society for Nondestructive Testing, Columbus
24. Eisberg R, Resnick R (1985) Quantum physics of atoms, molecules, solids, nuclei, and particles. Wiley, New York (ISBN-13: 978-0471873730)
25. Berger MJ, Hubbell JH, Seltzer SM, Chang J, Coursey JS, Sukumar R, Zucker DS (2008) NIST standard reference database 8 (XGAM). Online im WWW unter URL: http://physics.nist.gov/PhysRefData/Xcom/Text/XCOM.html. Accessed 08 July 2014
26. VDA Unterausschuss ZfP und DGZfP Fachausschuss Durchstrahlungsprüfung (editor): VDA-Prüfblatt 236-101/DGZfP Merkblatt D6: Anforderungen und Rahmenbedingungen für den Einsatz der Röntgencomputertomographie in der Automobilindustrie. 2008-11. Berlin : DGZfP (2008)
27. Partridge M, Hesse B-M, Müller L (2002) A performance comparison of direct- and indirect-detection flat-panel imagers. Nucl Instrum Methods Phys Res A 484:351–363

28. Hoheisel M (1998) Amorphous silicon X-ray detectors. J Non-Cryst Solids 227–230
29. Lee HR, Ellingson W (2001) Characterization of a flat panel amorphous Si detector for CT. J X-ray Sci Technol 9:43–53
30. Gondrom S, Schröpfer A (1999) Digital computed laminography and tomosynthesis-functional principles and industrial applications. In: Proceedings of international symposium on computerized tomography for industrial applications and image processing in radiology (DGZfP-Berichtsband 67-CD), DGZfP, 5
31. Simon M, Sauerwein C, Tiseanu I (2004) Extended 3D CT method for the inspection of large components. In: Proceedings of 16th world conference on nondestructive testing, WCNDT, 6 Sept 2004
32. Feldkamp LA, Davis LC, Kress JW (1984) Practical cone beam algorithm. J Opt Soc Am 6:612–619
33. Reiter M, Heinzl C, Salaberger D, Weiss D, Kastner J (2010) Study on parameter vatiation of an industrial computed tomography simulation tool concerning dimensional measurement deviations. Moscow ECNDT 2010
34. Smith BD (1985) Image reconstruction from cone-beam projections: necessary and sufficient conditions and reconstruction methods. IEEE Trans on Med Image MI-4(1):14–25
35. Kastner J, Harrer B, Requena G, Brunke O (2010) A comparative study of high resolution cone beam X-ray tomography and synchrotron tomography applied to Fe- and Al-alloys. NDT&E Int 43(7):599–605
36. Kastner J (2012) Habilitation thesis
37. Qureshi H, Malik M, Ahmad MA, Heinzl C (2012) Benchmarking of de-noising techniques for streaking artifacts in industrial 3DXCT scan data. Proc WSCG 2012:9
38. Otsu N (1975) A threshold selection method from gray-level histograms. Automatica 11 (285–296):23–27
39. Kastner J, Harrer B, Degischer H-P (2011) High resolution cone beam x-ray computed tomography for 3D-microstructure characterization of Al-alloys. Mater Charact 62:99–107
40. Kastner J, Zaunschirm S, Baumgartner S, Requena G, Pinto H, Garces G (2014) 3D-microstructure characterization of thermomechanically treated Mg-alloys by high resolution X-ray computed tomography. In: Proceedings ECNDT, Prague
41. Salaberger D, Kannappan KA, Kastner J, Reussner J, Auinger T (2011) Evaluation of computed tomography data from fibre reinforced polymers to determine fibre length distribution. Int Polym Proc 27(3):283–291
42. Weissenböck J, Amirkhanov A, Li W, Reh A, Amirkhanov A, Gröller E, Kastner J, Heinzl C (2014) FiberScout: an interactive tool for exploring and analyzing fiber reinforced polymers. IEEE Pac Vis Symp (PacificVis) 2014:153–160
43. Kastner J, Plank B, Salaberger D, Sekelja J (2010) Porosity measurement of carbon fibre-reinforced polymers by X-ray computed tomography. In: 2nd international symposium on NDT in aerospace, GE, Hamburg, 22–24 Nov 2010
44. Reh A, Plank B, Kastner J, Gröller E, Heinzl C (2012) Porosity maps—interactive exploration and visual analysis of porosity in carbon fiber reinforced polymers. Comput Graph Forum 31:1185–1194
45. Reh A, Gusenbauer C, Kastner J, Gröller E, Heinzl C (2013) MObjects—a novel method for the visualization and interactive exploration of defects in industrial XCT data. IEEE Trans Vis Comput Graph (TVCG) 19(12):2906–2915
46. Kastner J, Kickinger R, Salaberger D (2011) High resolution X-ray computed tomography for 3D-microstructure characterisation of a cellulose particle filled polymer foam. J Cell Plast 47:567–578

Chapter 9
Defect Inspection for Curved Surface with Highly Specular Reflection

Zhong Zhang and Changjiang Li

Abstract Highly specular reflection (HSR) curved surfaces and their inspection in most manufacturing processes mainly depends on human inspectors whose performance is generally subjective, variable, and therefore inadequate. An automatic vision inspection system offers objectivity, better reliability, and repeatability and is able to carry out defect measurement to evaluate the industrial part's quality. Thus, it is vital to develop an automatic vision system to monitor surface quality online. The main purpose of this chapter is to introduce a new defect inspection method capable of detecting defects on HSR curved surfaces, in particular, to create a complete vision inspection system for HSR curved surfaces (e.g., chrome-plated surfaces). In the first part of this chapter, reflection analysis of HSR curved surface is performed. And a new method is introduced to measure reflection properties of our inspection object. Then, a method is introduced to avoid the loss of defects and solve these challenges which result from various defects and complex surface topography on HSR curved surface. A set of images are captured under different illumination directions. A synthetic image is reconstructed from the set of images. The synthetic image appears more uniform intensity compared with the original image because those specular areas have been completely removed. Furthermore, all defects are integrated in the synthetic image. In particular, for more complicate curved surface, an improvement method is proposed and experiments also validate the method. Finally, a complete vision defect inspection system has been created. The lighting system with side and diffuse illumination is selected for our inspection system and it succeeds in reducing the specular reflection from a curved surface, although some brightness appears at the edge. System parameters and object pose are determined by comparing defect expressivity and specular ratio in the image.

Z. Zhang (✉)
Toyohashi University of Technology,
1-1 Hibarigaoka Tenpaku-cho, Toyohashi 441-8580, Japan
e-mail: zhang@is.me.tut.ac.jp

C. Li
Chongqing University of Science and Technology,
Chongqing University Town, Chongqing 401331, People's Republic of China
e-mail: lcjgdm@163.com

Moreover, all defects can be quickly extracted by combining template matching and morphology techniques. The presented automatic vision defect inspection system has been implemented and tested on a number of simulation images and actual parts consist of HSR curved surfaces.

Contents

9.1   Introduction

Products or parts with highly reflection specular surface, hereafter, we call highly specular reflection 'HSR,' are applied in kinds of fields due to their well appearance and surface properties. Surface appearance greatly affects the customer's perception of the product's quality and influences the decision to buy. On the other hand, a well-coated finish enhances the durability by protecting the product surface for corrosion. A painted surface such as automobile body, appliances such as washers, dryers, stoves, refrigerators, and cell phones out of the industrial paint shop require

rigorous quality inspection. In a car, there are many parts, such as cover-knob and car brand, are applied. These parts are chrome-plated surfaces and have strong specular reflection.

Inspecting objects with highly specular surface for scratches, cracks, or checking surfaces for proper finish are typical tasks of surface quality inspection in many industrial production processes [1–3]. Human experts are usually hired to perform the defect inspection. The results presented by individual human inspectors can be subjective and inconsistent due to unavoidable human errors. On the other hand, in order to completely inspect highly specular reflection surface, the human inspector necessarily varies the position of the test sample with respect to the illumination direct or his vision direction. In particular, for HSR curved surface, this operation can make defects become visible. So, the surface inspection tends to be time-consuming, labor intensive, and tedious or very difficult even for a trained operator though such an inspection method is effective. So, there is a long-standing desire in many industries to replace the current labor-based visual inspection of a product with machine-based inspection. The automatic computer vision inspection system plays a crucial role in this growing field [4], which offers objectivity, better reliability, and repeatability, and is able to carry out defect measurement to evaluate the industrial part's quality. For many years, researchers from both industrial and academic institutions have been conducting research on relevant topics and much machine vision systems have been developed to apply in varying fields. Kim and Koivo [5], for example, built a system for the automated inspection of wood surfaces. Piironen [6] developed an automated visual inspection system for rolled metal surfaces. Fernandez et al. [7] developed a surface inspection system for casting surfaces, Rohrmus [8] built a system for web defect detection, and Boukouvalas et al. [9] reported an inspection system for ceramic tiled surfaces.

However, for the HSR surfaces, methods enabling an automated inspection featuring a sufficient accuracy, reliability, and speed are still lacking. A few papers related to defect inspection on HSR surface have been published in the last two decades. A kind typical vision system has been built to inspect defects on mirror-like objects by using an array of point sources or lighted stripes. Nayar et al. [10] first proposed this structured illumination defect inspection technique. The technique uses an array of point sources to illuminate a specular surface. Defect inspection is performed by estimating features of the extended Gaussian image (EGI). Based on the technique, some new structured illumination techniques have been developed. For example, Caulier et al. [11–13] inspect specular surface by analyzing the deformation of light stripe. The illumination projects a structured pattern, which is reflected by the inspected surface and then projected/recorded by a camera. The generated stripe image depicts a periodical bright and dark stripe structure of period whose disturbances are synonymous of defective parts. Because of the rather heavy formalism involved in computing defect geometry from the deformation of the lighted stripes, these methods, although accurate, are extremely time-consuming. These applications solely enable the detection of the geometrical deformations of the inspected surface. The detection of other defects is not addressed. An alternative method to 3D triangulation is provided by deflect to

metric techniques, which are based on the measurement of the reflected light rays projected onto the object of inspection. Different shapes of specular object geometries can be recovered with micrometer resolutions by means of such a technique [14]. However, a preliminary calibration procedure of the recording setup (sensor, lighting) is necessary. The second necessitates at least two different illuminations techniques. The Parsytec company [15] has developed a dual sensor for recording object's surface with a diffuse and a direct light at the same time. Leon and Beyerer [16] proposed a technique where more than two images of the same object recorded with different lighting techniques can be fused in only one image. Aluze et al. [17] presented a vision system to detect bump defects (e.g., dust and hair) on mirror-like objects. The vision system comprises an innovative lighting solution to reveal defects onto highly reflective non-plane surfaces. Several image acquisitions are performed to build a synthetic image, where defects clearly appear white on a mid-gray background. The major disadvantage of those approaches is of course that they necessitate more than one illumination. The direct consequence is that their integration in the industrial process is more complex and that the data processing chain is more extensive.

In contrast to the above approach, another possible technique is to inspect defects from diffuse image. The first technique is to capture diffuse reflection from specular surface by special illumination source (e.g., diffuse illumination [18, 19]). In fact, this illumination is not efficient to highly specular surface although it can be applied to some metal surface defect inspection. Another technique is to obtain diffuse reflection components by means of removing or separate specular from captured image. For example, Zhang et. al [20, 21] presented a special random sampling method (SRSM) to separate the specular reflection components and convert the input images into synthetic images that contain only diffuse reflection components. Furthermore, in order to identify defects in HSR surfaces, a template matching method was used in their proposed inspection system. The major disadvantage of the method is that more than twenty images are captured and lead more complicate image processing algorithm and time-consuming. So, it is difficult to be applied in real-time defect inspection.

Another challenge on a HSR curve surface is that defects on the HSR curve surface with very weak diffuse reflection are too sensitive to illumination direction to form recognizable information. The reason is that curve surface topography is complex and defect properties such as shape, direction, and location are complex and various. Figure 9.1 shows the photograph of a chrome-plated part, which is taken in a natural environment, and presents the above appearances and challenges. The environment surrounding the object can be observed because the HSR surfaces of the mirror-like objects reflect light rays directly. In addition, some high light completely covers some surface, so one does not know whether there are defects in these bright regions. Based on the above discussion, one can conclude that not all defects on surface can be illuminated and form recognizable illuminated spots in one obtained image of the HSR curve surface.

In this chapter, in order to improve the above-mentioned problem, a new automatic vision defect inspection system for HSR curve surface has been
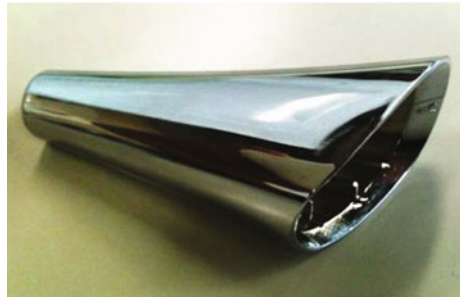
developed. Figure 9.2 shows the work chart of the inspection system. A set of images are captured under various illumination conditions while the object and viewpoint are fixed. By processing the set of images, the surface quality will be assessed referring to the corresponding guideline. This system is to achieve defect inspection on HSR curve surface (e.g., chrome-plated surface). It is well known that the type surface is not only with highly specular reflection while very weak diffuse reflection but also with circumscribed curve surface, and it is a significant advancement relative to previous method. In this chapter, a novel illumination solution and a pose configuration investigation are provided to ensure high defect expressive and considerable less high light reflection on inspection region. In addition, the proposed system captures a set of images illuminated, respectively, under different illumination directions to completely avoid the loss of defects and solved these challenge which result from various defects on HSR curve surface. Moreover, a synthetic image is reconstructed from the set of images in the gradient field based on the observation that specular reflection will appear at different location in the set of images. In the synthetic, all defects are integrated with high contrast and bright regions are eliminated. To the end, a novel defect inspection method, which combines the morphological technique and the template matching, is proposed to correctly identify kinds of defects from the synthetic image. Some experiments on HSR curve surface such as chrome-plated acrylonitrile butadiene styrene (hereafter referred to as chrome-plated ABS) resin were carried out using our system. Experiments show that this system can reliably detect defects on HSR
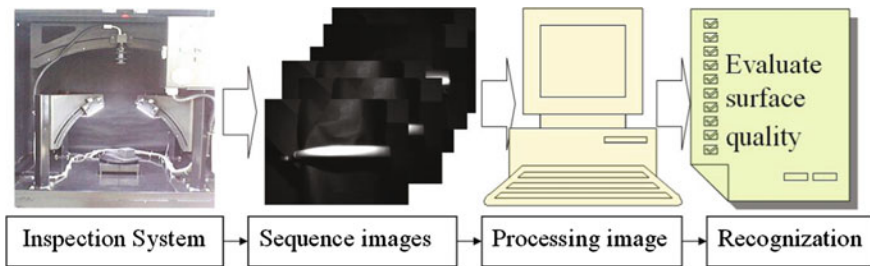
Fig. 9.2 The flowchart of the new inspection system

curve surface and it is robust to the shape, location of the defect which is extracted a synthetic image reconstructed from a set of images. Moreover, the inspection system is cost-efficient because of the low algorithm complexity and suitable for use in industrial applications. The primary significance and novelty of this work is in achieving defect inspection on HSR curve surface (e.g., chrome-plated surface). The object in this chapter is not only with highly specular reflection but also with circumscribed curve surface, it is a significantly advancement relative to previous method.

9.2 Theory Background Related to Defect Inspection

In this section, the theory background and some related work will be presented. Computer vision are used to analyze surface quality in this investigation; therefore, all the surface inspection technologies analyze information gathered from the light rays reflected from the inspection product surface to perform defect inspection. In Sect. 9.2.1, some of the most important results related to the modeling of surface physical reflectance are discussed. Torrance-Sparrow [22] and Ward [23] unified reflectance framework is well recognized as one of most accurate surface physical reflection models and provides the scientific based for this study. So, the two reflection models are specially introduced to analyze reflection how to vary from defect region to flawless region. These theories will be used to design defect inspection approach. Section 9.2.2 discusses some latest inspection methods and some current available online inspection systems on such specular surfaces.

9.2.1 Review of Surface Reflection Models

Some earlier approaches to surface physical reflection models [24–26] have assumed that the surfaces are Lambertian [27], in which the incident light is scattered by the surface evenly around the surface normal direction. However, an ability to understand specular features is valuable for any vision system, which must interpret images of glossy surfaces; for example, a specular highlight can yield additional information about the local curvature of the surface. So, it can be used to resolve convex or concave ambiguities.

Later representative research studies on this subject include Phong [28], who proposed a parameterized continuous function to represent specular reflectance and used it to create images of the objects; Ikeuchi [29] and Sanderson [30] used the double delta specular model to determine the shape of specular surfaces by a photometric stereo and structured highlight technique; Healey and Binford [31] used the Torrance-Sparrow model [22] to analyze monocular images to estimate local shape curvature. In addition, Nayar [32] proposes a more accurate unified

reflectance framework for the machine vision that predicts the reflectance of monochromatic light from both smooth and rough surfaces. It is based on both geometrical and physical reflectance models.

In general, the reflection from a surface consists of three primary reflectance components: the specular lobe I_{sl}, the specular spike I_{ss}, and the diffuse lobe I_{dl}, as shown in Fig. 9.3 and the coordinate system used for the model is shown in Fig. 9.4. The radiance of the surface in the sensor direction may be expressed as the sum of all three components, as shown in Eq. (9.1):

$$I = I_{dl} + I_{sl} + I_{ss}, \tag{9.1}$$

where each reflectance component can be shown as follows:

1. Diffuse lobe I_{dl}: The Lambertian model is used to represent the diffuse lobe component, for which the magnitude is distributed evenly for all the viewing directions. The diffuse component is represented by the Lambertian model, as shown in Eq. (9.2)

$$I_{dl} = K_{dl}\cos(\theta_i), \tag{9.2}$$

where K_{dl} denotes the strength of the diffuse lobe, and θ_i is the angle of the incident light rays.
2. Specular lobe I_{sl}: A Torrance-Sparrow model [22] is used to predict the Specular lobe:

$$I_{sl} = K_{sl}\exp\left(-\frac{\alpha^2}{2\sigma^2}\right)F(\theta_i', \eta')G(\theta_i, \theta_r, \phi_r), \tag{9.3}$$

where K_{sl} is the magnitude of the specular lobe, $\exp(-\alpha^2/2\sigma^2)$ is an exponential term that describes the slope distribution of the micro-facets assuming a normal

Fig. 9.3 Reflection components [32]
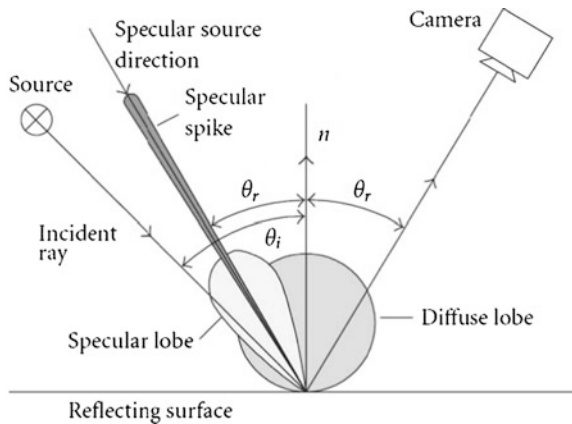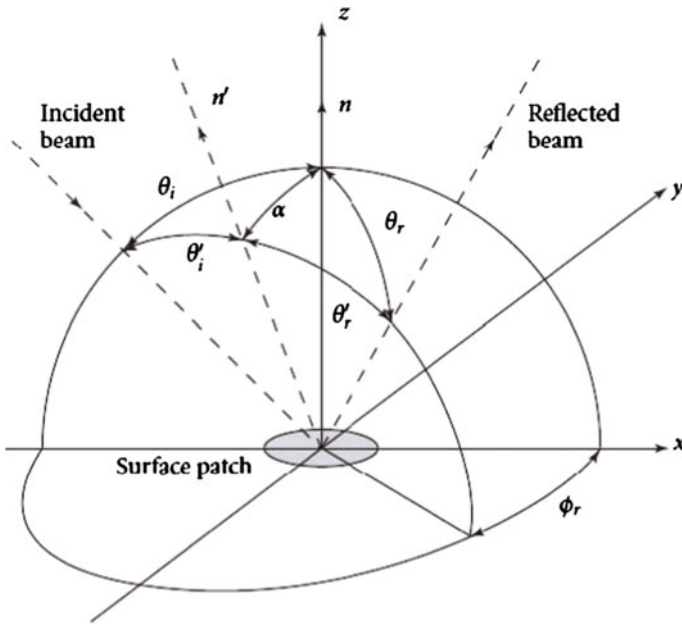
Fig. 9.4 Coordinate system for the reflection model [32]

distribution, α is the angle between the normal vector of the micro-facet n' and the normal vector of the macro-surface n, that is $\arccos(n \times n')$, and θ is the standard deviation of the surface height. Smaller values of σ indicate the surface is smoother ($\sigma = 0$ represents a perfectly smooth surface, like a pure mirror). Larger values of σ imply the surface is rougher. $F(\theta_i, \eta')$, where θ_i' is the local angle of incidence and η' is the complex index of refraction and is the Fresnel coefficient, and under certain conditions it can be defined as a constant value. $G(\theta_i, \theta_r, \phi_r)$ is a geometric attenuation factor, and it describes the shadowing and masking effects of micro-facets by adjacent micro-facets. When the viewing direction coincides with the mean normal vector n, the factor G equals unity over a large range of θ_i.

3. Specular spike I_{ss}: The specular spike component is a very sharp function, which is concentrated in a small region around the specular, or mirror, direction. It is approximately described by a delta function.

$$I_{ss} = K_{ss}\delta(\theta_i, \theta_r)(\theta_r), \tag{9.4}$$

where K_{ss} is the strength of the specular spike component. The symbol δ represents the delta function.

9.2.2 Surface Reflection Analysis on Highly Specular Surface

In this subsection, the specular reflection from highly specular surface will be further discussed to develop the following defect inspection approach. Reflection off metallic surfaces is different from that off inhomogeneous surfaces. In case of an inhomogeneous (dielectric) material, the following four reflected light elements shown in Fig. 9.5 have been considered [20]: (1) Specular reflection that reflects once off a surface, which is flatter than the wavelength of the incident light. (2) Diffuse reflection that has reflected two or more times off a rough surface, which consists of material smaller in size than the wavelength of the incident light. (3) Diffuse reflection that diffracted on the surface, which consists of material not greater in size than the wavelength of the incident light. Usually, the reflected light element can almost be disregarded because it is very small except when the surface of the object has a periodic structure of order roughly equivalent to the wavelength of the light. (4) Diffuse reflection that first penetrated the surface and then escaped the surface after a repetition of reflections off colored matter (pigment) on the inside of the dielectric material.

In the reflected light off a metallic material, there is no fourth element. This is because the metallic material has conductive electrons, so the electric field of the light is covered by Dolude's Law and the light cannot enter the metal. Furthermore, refection off a metallic surface is different from that off a dielectric surface. For one thing, the reflections have different polarization properties. For unpolarized incident light, the reflected light off a metallic surface is still unpolarized, while the reflected light off an inhomogeneous surface is polarized. A second reason is that for smooth metallic surfaces, specular reflection is very strong and diffuse reflection is too weak. Moreover, the surroundings are often observed as specular reflections.

Figure 9.6 shows the 3D surface of a scratch. The flaws degrade the surface smoothness. The standard deviation of the approximately normal distributed surface slopes θ_1 is for the errors lower than for the intact surface assumed as θ_2. The change of the surface height in case of a flaw is at least twice that of the flawless surface, and also the frequency of the main oscillation of the roughness profile is
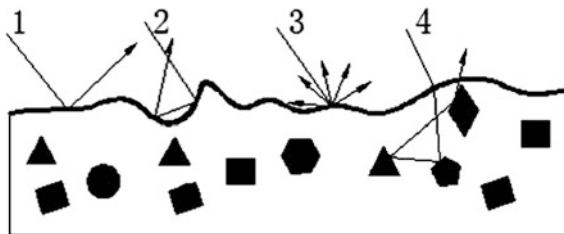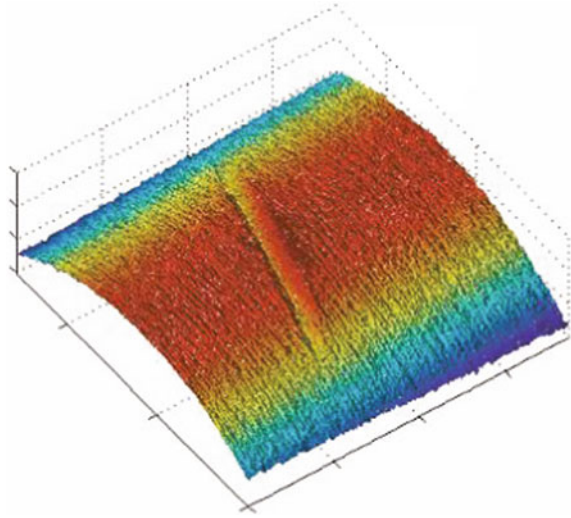

**Fig. 9.5** Reflection of light off a dielectric material

Fig. 9.6 3D profile of a scratch [64]

lower, and the slopes are smoother, and these effects result in a lower θ. This difference causes a shift of the reflection property.

The reflection intensity of the specular lobe component is computed according to the model [22] and plotted in Fig. 9.7. The viewing direction is fixed at $\theta_r = 0$ and the source direction θ_i is varied. Figure 9.7 shows the relative reflected intensity for a flawless region and a defect region is computed according to the reflection model [22] which is restricted to the specular lobe component. It is shown in Fig. 9.7 that the difference in the reflected intensities for the intact surface and defect surface reflects the difference in surface roughness. The largest difference is given for an

Fig. 9.7 The relative reflected intensity for a flawless region and a defect region, *dash* present flawless region and *solid line* defect region
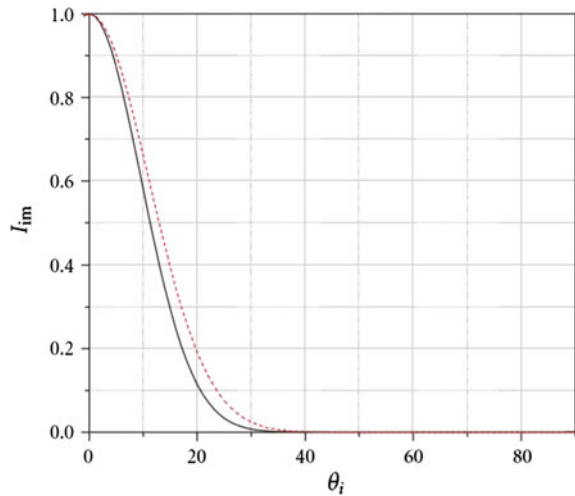
angle of illumination θ_i in the range from 100 to 300. Hence, an illumination angle in this range should lead to an optimum contrast.

On the other hand, the specular spike component is a very sharp function, which is concentrated in a small region around the specular, or mirror, direction. It is described by a delta function. Figure 9.8 shows two examples to present the above phenomena. As shown in Fig. 9.8a, the observation angle θ_r is much bigger than the illumination angle θ_i. So, specular spike and lobe from a flawless surface cannot be captured by camera. On the other hand, for highly specular surface, its diffuse reflection is very weak close to zero. According to above two aspect analyses, a very dark image will be obtained. In contrast, Fig. 9.8b shows a plane with two defects, in which one is a dent and another is a bump. It is very obvious that normal vector of defect distinctly varies from the perfect region. It is not difficult to understand that the bright spot present the specular spike from the defect region. It is because that the defect region change the normal vector of the surface, which result in the reflection from defect region enter the vision sensor. Of course, one cannot confirm that it definitely results from specular spike. One can conclude the following conclusions from the example. First, under some conditions which involve view sensor, illumination and so on, the reflection from defect region will distinctly different from flawless surface. This difference is very enormous for

Fig. 9.8 Examples of the flawless specular surface, where **a** shows a flawless specular surface and **b** shows a plane with two defects
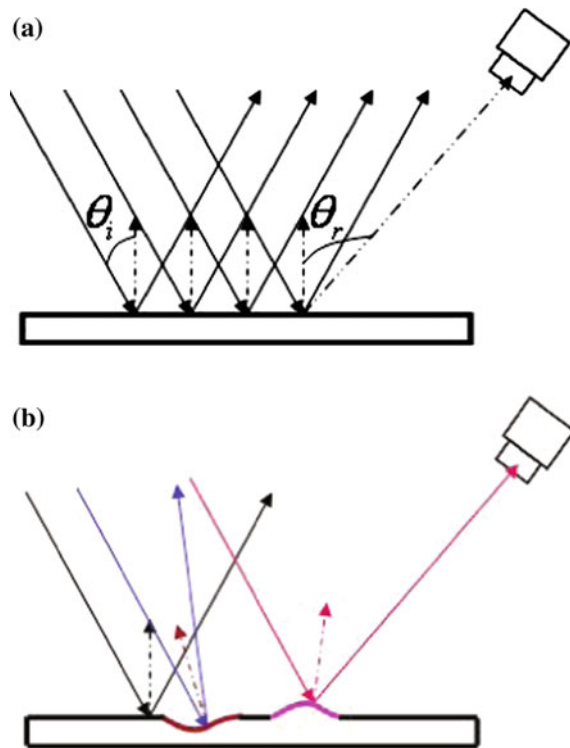
highly specular surface. Of course, this case also is very significant to defect inspection. Second, in contrast, some defects can be found under the same condition. This problem will be discussed and solved in the following chapter.

9.2.3 Summary

In this section, the theory background related to defect inspection on highly specular reflection surface while very weak diffuse reflection was presented. The following conclusions can be obtained.

1. Specular lobe of defect region will be significantly different from flawless region due to the different roughness θ. The difference is in relation to the illumination angle θ_i. The specular spike is a mirror-like reflection. It only can be captured when the illumination angle θ_i is equal to the view angle θ_r.
2. Under some conditions which involve view sensor, illumination and so on, the reflection from defect region will distinctly different from flawless surface. This difference is very enormous for highly specular reflection surface while very weak diffuse reflection. So, the reflection from defect region will distinctly different from flawless surface. Of course, this case also is very significant to defect inspection. In contrast, some defects cannot be found under the same condition. This problem will be discussed and solved in the following chapter. So, defect inspection can be performed by means of the phenomena, and it is a relatively simple and efficient approach, especially for highly specular reflection surface while very weak diffuse reflection.

9.3 A Numerical Simulation for Defect Inspection on Highly Specular Surface

In this section, we will present a numerical simulation to demonstrate the some reflection phenomenon from various defects on smooth, highly specular surface (i.e., very glossy). Numerical simulation is a flexible, practical tool for efficiently investigating machine vision system hardware and software design for myriad industrial applications. Simulated images can aid in both image understanding and vision system design, significantly reducing the engineering time to design and successfully implement a vision system. Using a reflectance model for isotropic, opaque surfaces, we will utilize the properties of smooth, specular coatings to investigate the simulation image different between flaw surface and perfect surface. Moreover, we especially present how do reflection change from various defects (i.e., defect shape, defect direction).

9.3.1   *Image Generation Approach*

Given the 3D surface geometry, illumination direction and surface reflectivity, an image is rendered to simulate the camera process with respect to simulating the image formation process. The observed (measured/captured) intensity of an image depends on four main factors: (1) illumination, which is defined by its position, direction, and spectral energy distribution (omnidirectional, spotlight …). (2) Surface reflectivity of the object, which is known as albedo, it entails the information about how the object reacts with light in terms of the amount of incident light being reflected; in computer graphics literature, it is denoted as surface reflectance coefficients or more commonly as surface material. (3) Surface model, which describes the surface geometry of an object. (4) Camera capturing the object, which is defined by its intrinsic and extrinsic parameters.

9.3.1.1   Surface Model

There are many possible representations, for example, height maps and meshes. In this subsection, the height maps [33] is briefly introduced. The shape representation used most frequently is height maps. A height map stores 3D shape as a set of height or depth values defined over a plane (See Fig. 9.9). Mathematically, a height map can be specified as the set of points $(x, y, z) = f(x, y)$ for a function $f(\cdot)$. For example, Eqs. (9.5), (9.6) are the explicit surface equations of a vase. Height maps are very natural for computer vision applications, since all points in the world are projected on the image plane meaning there is at most a single visible surface point per pixel in an image.

$$Z(x, y) = \sqrt{f^2(y) + x^2}. \tag{9.5}$$

Fig. 9.9 The 3D plot of a vase
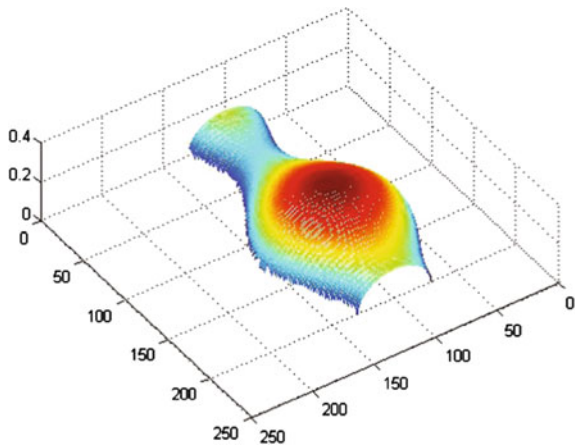
$$f(y) = 0.15 - 0.1y(6y + 1)^2(y - 1)^2(3y - 2)^2,$$
$$0.5 \leq x \leq 0.5, 0.0 \leq y \leq 1.0 \tag{9.6}$$

If the height map is continuous and differentiable, that is, if $z_x = \frac{\partial z}{\partial x}$ and $z_y = \frac{\partial z}{\partial y}$ exist for all (x, y), then the gradient is defined as (p, q), where $p = z_x$ and $q = z_y$. Moreover, the surface normal map can be defined in terms of the gradient,

$$n = \frac{(-p, -q, 1)^T}{\sqrt{1 + p^2 + q^2}}. \tag{9.7}$$

9.3.1.2 Camera Model

Before starting to analyze an image, it is necessary to understand the basic fundamentals involved in image formation. The camera model describes a way of relating the real Cartesian coordinates of the position of an object located in real space to its location in the discrete pixel space of the image pixel array [34]. Figure 9.10 illustrates the basic geometry of the camera model. (x_w, y_w, z_w) is the coordinate of the object point P in the 3D world coordinate system. (x, y, z) is the 3D coordinate of the object point P in the 3D camera coordinate system, which is

Fig. 9.10 Definitions of reflection parameters and angles
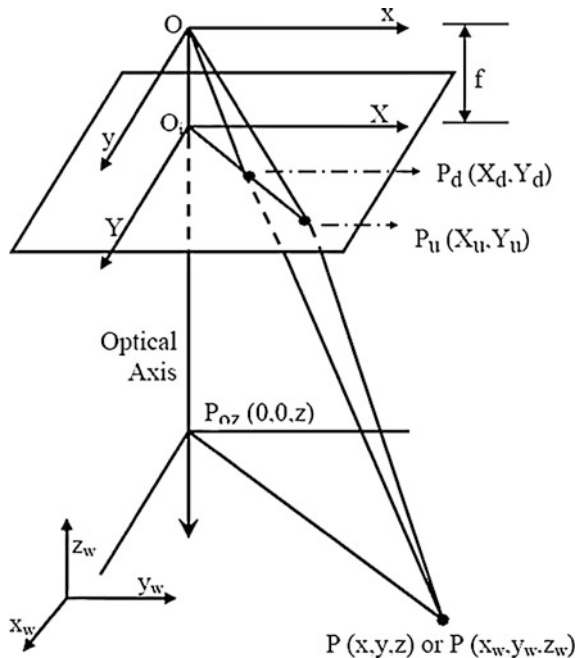
centered at the point O, the optical center, with the z-axis the same as the optical axis. (X, Y) is the image coordinate system centered at the intersection of the optical axis z, with the front image plane at O_i and parallel to x and y axes. f is the distance between front image plane and the optical center (O). (X_u, Y_u) is the image coordinate of (x, y, z) if a perfect pinhole camera model is used. (X_d, Y_d) is the actual image coordinate which differs from (X_u, Y_u) due to lens distortion. However, since the unit for (X_f, Y_f), the coordinates used in computer, is the number of pixels for discrete image in frame memory, additional parameters need to be specified and calibrated that relates the image coordinate in the front image place to the computer image coordinate system. The overall transformation from the (x_w, y_w, z_w) to (X_f, Y_f) is depicted in Fig. 9.11. It is essential to calibrate the camera in order to be able to relate the computer frame memory coordinates to real-world coordinates. This ability will be the key to the effective characterization of defects presented in Sect. 9.5.

Fig. 9.11 Flowchart of world coordinates to computer frame memory coordinates

(x_w, y_w, z_w) 3-D world coordinate

Rigid body transformation from (x_w, y_w, z_w) to (x, y, z), Parameters to be calibrated: R, T

(x, y, z) 3-D camera coordinate system

Perspective projection with pinhole Geometry, Parameters to be calibrated: f

(Xu, Yu) Ideal undistorted coordinate

Radial lens distortion, Parameters to be calibrated: K

(X_d, Y_d) Distorted image coordinate

Parameters to be calibrated: uncertainty scale factor for image X calibrate

(X_f, Y_f) computer image coordinate in frame memory

9.3.2 An Efficient Reflection Properties Measurement for Glossy Surface

In this subsection, we present an efficient photometric stereo method to estimate robustly BRDF parameters acquisition without the 3D geometric model of object. We estimate reflectance from a set of monochrome images of the surface of interest. Metallic surface has homogeneous reflectance properties, and its diffuse and specular reflectance remain constant over the whole surface. So, in order to reduce the effect of noise and shadow, a best fitting window in one of the input images is defined. All points on the surface are classified according to their radiance values into the diffuse points and the mix ones. Two-step optimization is adopted to estimate all BRDF parameters. In the first step, the diffuse reflectance and surface normal are recovered using these diffuse points. And then specular reflectance and roughness of Ward [23] model can be recovered because its normal and diffuse reflectance are known. In the second step, a robust iteration process is carried out to estimate BRDF parameters in a best window of image.

9.3.2.1 Reflection Model

In this subsection, we give a brief overview of the reflection model used in our method. Generally, reflection models are described by linear combinations of two reflection components: the diffuse and the specular reflections. The diffuse reflection component represents reflected rays arising from internal scattering inside the surface medium. The specular reflection component, on the other hand, represents light rays immediately reflected on the object surface. The following assumptions are proposed: (1) The object is far from the camera, therefore the camera projection can be approximated by an orthographic projection; (2) The scene is illuminated by a point light source located far away from the surface. The Ward BRDF [23] is used to model the reflectance property of an object's glossy surface. The model has been validated by collecting many measurements from real images. The expression for Ward reflectance model is given by:

$$f(i,o) = \frac{\rho_d}{\pi} + \rho_s \frac{1}{4\pi\sigma^2 \sqrt{\cos\theta_i}\sqrt{\cos\theta_o}} e^{-\frac{\tan^2\delta}{\sigma^2}}, \qquad (9.8)$$

where i and o denote the incident and outgoing light directions. θ_i and θ_o are polar angles of incident and view directions, respectively, where ρ_d is the diffuse albedo and it determines the proportion of incoming light reflected diffusely. The higher the value of ρ_d, the brighter the surface. The specular albedo ρ_s controls the proportion of incoming light that is reflected specularly. Small values of this

Fig. 9.12 Definitions of reflection parameters and angles
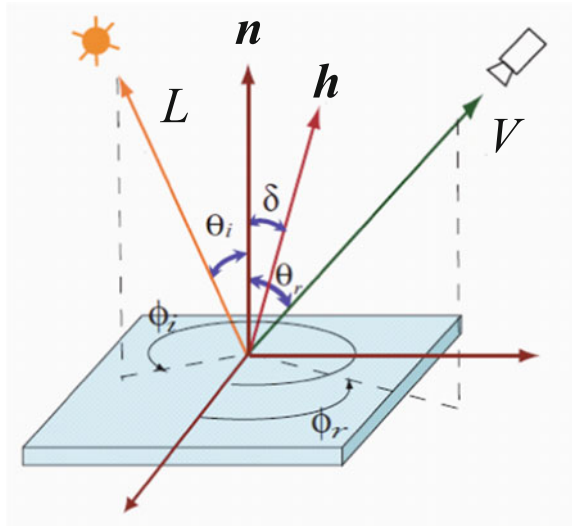
parameter yield matte surfaces while higher values yield glossy and metallic surfaces. The parameter σ is the standard deviation of the surface roughness at a microscopic scale, which determines the size of the specular lobe on object surface. Changing this parameter leads to changes in the 'spread' of the specular reflection. Small values of the roughness parameter lead to crisp specular reflections, while large values lead to blurred reflections such as unpolished metals. The angle δ shown in Fig. 9.12 is the angle between the surface normal $\mathbf{n}$ and a halfway vector $\mathbf{h} = [h_x, h_y, h_z]^T = \frac{\mathbf{s+v}}{|\mathbf{s+v}|}$ [23].
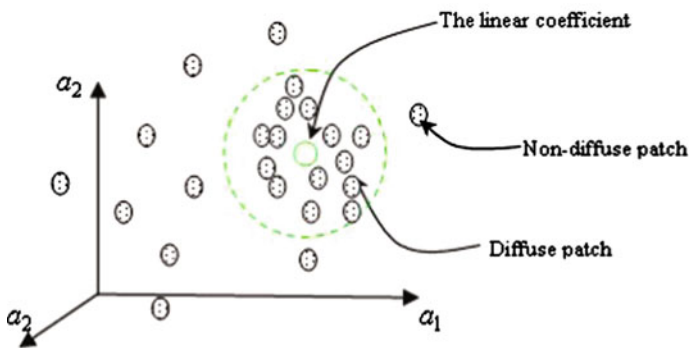
Fig. 9.13 Definitions of reflection parameters

9.3.2.2   Classify of Reflection Component

1. Calculation of coefficients vector

   If both the target objects and the viewing position are fixed, and the Lambertian reflection model is assumed, the observed image depends only on the lighting direction. In other words, linear dependence of the illuminant vectors leads to the same linear Eq. (3.5) for the corresponding pixel intensities, if the Lambertian assumption holds. Literature [35–37] shows that if a single point light source is assumed at infinity, any four images $I_k(k = 1, 2, 3, 4)$, illuminated by four illumination vectors $L_k(k = 1, 2, 3, 4)$, the following relation holds.

$$a_1 I_1 + a_2 I_2 + a_3 I_3 = I_4, \tag{9.9}$$

   where $\vec{a} = [a_1, a_2, a_3]^T$ is called a linear coefficients vector. In fact, any four vectors are linearly dependent in the three-dimensional world. So, the coefficients vector $\vec{a}$ could be computed directly from the known illumination vectors. The following relation between the linear coefficient and the illumination vectors L_k which corresponds to four images $I_k(k = 1, 2, 3, 4)$, respectively, is shown as follows:

$$a_1 L_1 + a_2 L_2 + a_3 L_3 = L_4, \tag{9.10}$$

2. Classification of reflection components

   Unfortunately, in practice, real image does not satisfy Eq. (9.9), because shadows and specular reflection components are observed. Let us consider a surface patch p with albedo ρ_d and normal **n**. Under a 4-source image configuration (and for gray images), this patch gives rise to four pixel values $I_k^p(k = 1, 2, 3, 4)$. For the intensity value of pixel in the fourth image of the patch p, the corresponding coefficients vector a^p can be obtained. If in the four pixel values $I_k^p(k = 1, 2, 3, 4)$, shadows, and specular reflection components are not are observed, the a^p should be equal the coefficients vector **a** obtained according to Eq. (9.10). On the other hand, if in one of the four values there are specular or shadow components, the a^p will be isolated from the correct vector **a**, Fig. 9.13 illustrates the basic idea of the method, where d^p is set as the Euler distance between a^p with **a**. Therefore, for each surface patch p, we detect the d^p value and classify them into diffuse patch and mix one. Let us set a radius T_l, if the $d^p \leq T_l$, the patch p is considered as diffuse patch. If $d^p \geq T_l$, the patch p is considered as mix one. It means that at least three of the four pixel values do not include shadow or specular reflection components.
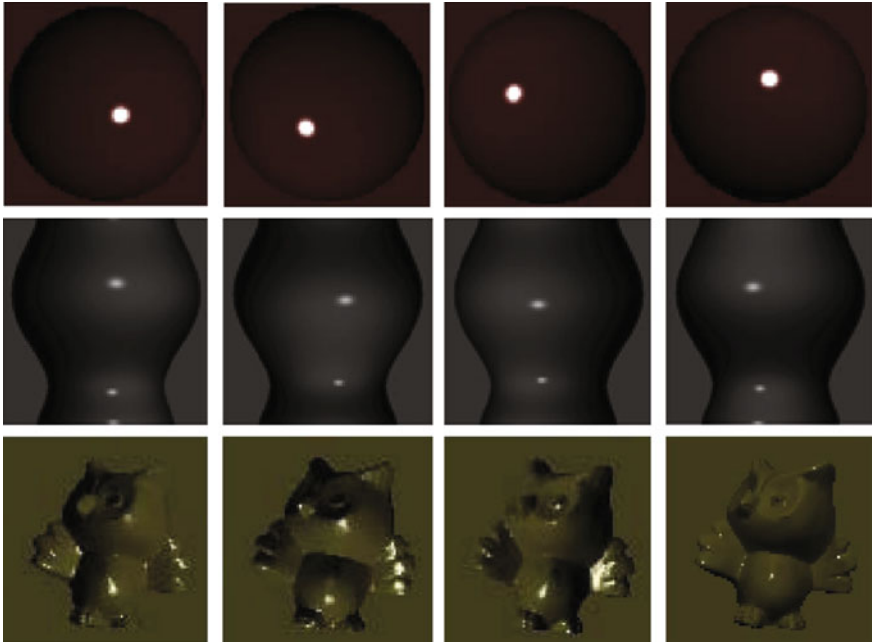
Fig. 9.14 Three group images of three different objects

9.3.2.3 Estimation of Reflectance Parameters

In this subsection, the estimation method with respect to reflectance of Ward reflection model for glossy surface is presented in detail. One simplifies the problem at this stage by assuming that the surface under observation has homogeneous reflectance properties and that its geometry is unknown. We refer to ρ_d as the diffuse parameter and ρ_s, and α as the specular reflection parameter.

1. Define the best window
   The estimation of reflectance parameters of many of the past studies is based on global approach, that is, analysis of all data in the image. The global approach is easy to be affected by noise, saturation, and shadow pixel. In addition, calculation is also large. However, specular reflection in the statistical is partial and concentrated. For our object, the reflectance parameters are consistent. So, this makes the analysis of the local window can become possible. Under the same conditions, specular pixels are brighter than surrounding ones, therefore the best analysis window is we determined according to the brightness of the image information. The average brightness I_1 value of the whole image is calculated. Particularly, when the background brightness is relatively low value of many

pixels, the average brightness I_{avel}^2 among those pixels which intensity is higher than I_{avel}^1 is calculated again. For color images, the intensity is computed by the equation $I = [0.3086, 0.6094, 0.0820] \times [\text{RGB}]^T$. Without loss of generality, the size of window is set as 1/16 of the size of image. Then, the window Ω is determined as best fitting window in which the total number of pixel with higher intensity than I_{avel}^2 is the largest. The follow-up process will be done in the best window Ω.

2. Calculate initial reflectance

The surface normal and the initial reflectance are calculated. The first term of (9.8) is extracted, and the diffuse components of the patch p can be modeled as Eq. (9.11).

$$I_{pdi} = \rho_d L_i n_p^T, \tag{9.11}$$

where I_{pdi} is the pixel intensity of patch p with the albedo ρ_d and the unit surface normal n_p, L_i is the ith lighting direction (a unit vector). Here, the pixel intensities are stacked to obtain the pixel intensity vector $\mathbf{I}_{pd} = [I_{pd1}, I_{pd2}, I_{pd3}, I_{pd4}]^T$. In addition, the illumination vectors are stacked to form the illumination matrix $[L] = (L_1, L_2, L_3, L_4)^T$ (the square bracket is used to denote matrices). So, the intensities of the obtained (grayscale) pixels can be expressed as Eq. (9.12):

$$I_{pd} = \rho_d [L] n_p, \tag{9.12}$$

For mix patch in the region Ω, the pixels which intensity value is zero or one are considered as shadows and saturated pixels, respectively. So, shadows and saturated pixels are got rid of and get the pixel collection Ω'. Now, reflectance for those remained mix patch in the region is only estimated. So, the pixels contain specular components can be considered; it is obvious that only the three pixels from the four ones can be used. So, the darkest three intensity values and their corresponding light direction are used to form $I_p d$, $[L]$ to calculate the $\rho_d n_p$. For diffuse patch in the region Ω, there are more than three input images, and the illumination matrix $[L]$ is not square and cannot be inverted directly. Following the standard least square error procedure, the albedo ρ_d and the normal n can be recovered by means of Eq. (9.13).

$$\rho_d n_p = ([L]^T [L]^{-1})[L]^T I_{pd}, \tag{9.13}$$

So, the direction of the obtained vector $\rho_d n_p$ can considered as the surface normal n_p. On the other hand, the initial diffuse parameter ρ_d^* is determined by calculating an average of the length of $\rho_d n_p$ in the region Ω. To obtain the initial specular

parameters, the second term of Eq. (9.8) is extracted, and the specular components of the patch p can be modeled as Eq. (9.14).

$$I_\psi = \rho_s \frac{1}{4\pi\sigma^2} \sqrt{\frac{\cos\theta_i}{\cos\theta_o}} e^{\frac{\tan^2\delta}{\sigma^2}}, \tag{9.14}$$

If the normal of one surface point is known and the corresponding θ_o, θ_i and β are determined, then the formulation of I_ψ can be represented by two unknown parameter ρ_s, σ. On the other hand, the specular components of the input image can be obtained by subtracting the aforementioned diffuse estimated from the input image. Then, the initial specular parameters ρ_s^*, σ^* are estimated by minimizing the following Eq. (9.15).

$$E = \sum_p (I_{pi} - I_{pi}^* - I_{psi}), \tag{9.15}$$

where I_{pi} is the pixel intensity of patch p of the ith image. To obtain desirable reflectance, all reflectance parameters together are optimized by minimizing Eq. (9.16):

$$\text{argmin} \sum_{p\in\Omega'} (I_{pi} - I_{pi}^*), \tag{9.16}$$

where I_{pi} is represented by Eq. (9.8). The initial values of reflection parameters were obtained in the aforementioned initial reflectance estimate. Since the initial value is close to optimal value, the whole process is guaranteed to converge to an optimal solution and the computer cost is shorter than the convenient approach. The above optimization process shows that the reflectance parameters can be efficiently estimated.

Here, we investigate some result of the proposed method. Three groups of images of objects, which are hemisphere, vase, cat, respectively, are shown in Fig. 9.14, and their corresponding reflectance property is shown in Table 3.1. For the color image, the same computation is extended independently for each color band. Every group includes four images illuminated with varying the angle θ_i in increments of 90° from 0° to 360°. The experiment results are shown in Table 9.1. In Table 9.1, ρ_d, ρ_s, σ, and $\rho_d', \rho_s', \sigma'$ are the original reflectance and recovery ones, respectively.

Table 9.1 The evaluation reflectance parameters

$\rho_d(r,g,b)$	$\rho_d'(r,g,b)$	(ρ_s,σ)	(ρ_s',σ')
(0.63, 0.09, 0.09)	(0.63, 0.09, 0.09)	(0.10, 0.05)	(0.10, 0.05)
(0.56, 0.56, 0.56)	(0.56, 0.56, 0.56)	(0.30, 0.10)	(0.30, 0.10)
(0.43, 0.41, 0.18)	(0.43, 0.41, 0.18)	(0.40, 0.15)	(0.40, 0.15)

9.3.3   A Numerical Investigation for Defect Detection
on Specular Surface

According to a DU PONT, there are fourteen types of typical paint defects that
affect the quality of surface appearance. The defects detected in this work are seed
defects (areas noticeably higher than the nominal surface) and spot defects (areas
noticeably below nominal height). Their geometries have dimensions ranging from
approximately a millimeter to several centimeters in height and lateral extent;
smaller defects are not consistently discernible by professional human inspectors.
Seeds can be characterized as raised bumps with a discernible height which are
present on paint surfaces. They can be caused by the thinness of the color coat;
improper filtration, which may results in contaminants trapped in the coating; out of
spec raw materials or the incorrect reducing solvent during the painting process.
Spots are planar: spot-like or linear faults such as scratches. They have little geo-
metric variation, but a relatively intense contrast to the non-defective surrounding
surface.

9.3.3.1   The Defect Type and Mathematical Model

Defects are zones where the surface is affected by geometrical imperfections. The
surface is still smooth in the defect zones, so the specular property is conserved.
Defect's shapes are irregular and their sizes are various, ranging from millimeters to
centimeters. So, it is challenging to find a single mathematical model that can
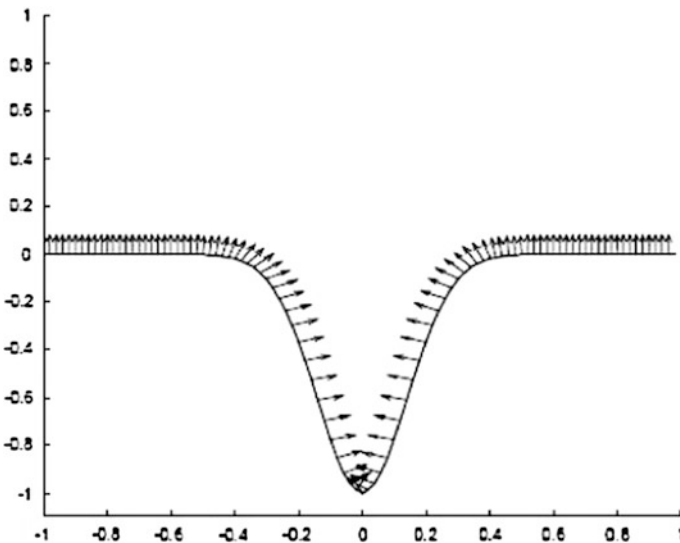
Fig. 9.15 Dent defect model and normal vectors

accurately and completely describe the shapes of all defects. However, a reasonably simplified mathematical model can still aid in characterizing a defect's geometric properties by facilitating an understanding and qualitative analysis of the surface specular reflection around the defect areas; such a model can also aid in determining the defect inspection mechanism. The Gaussian function characterizes the smoothness of the surface. So, a dent or bump can be modeled by a standard Gaussian function (Eq. (9.17), its Gaussian curved in Fig. 9.15). Normal vectors are computed from the surface model

$$\phi(x) = \frac{1}{\sqrt{12\pi\sigma}} e^{\left(-\frac{x^2}{2\sigma^2}\right)}. \tag{9.17}$$

9.3.3.2   The Numerical Simulation of Defect on Specular Surface

A perfectly specular surface is defined as a surface that reflects a light ray in a single direction from any given point reflected on it, where the angle of the reflected ray with respect to the surface normal at the point is equal to the angle between the normal and the incident ray. The normal, incident, and reflected rays lie in the same plane. Specular reflectance models are widely used to describe mirror-like reflections from specular surfaces such as glass, ceramic, polished metal, and some plastics; the reflected illumination from these real-world materials is concentrated in the specular direction. So for a defect-free highly specular flat surface, a camera at an off-specular angle will capture essentially no light energy. This is illustrated in Fig. 9.8a, where the camera angle θ_r is significantly greater than the incident angle θ_i. However, when any geometric imperfection exists, such as a bump defect present on the inspected surface, those incident rays that hit the defect surface will yield local specular reflection rays whose directions are significantly different from those produced by the background surface, as illustrated in Fig. 9.8b. Because the curved of the bump defect surface is continuous, the resulting specular reflections will be distributed in a large range of angles as shown. In this situation, although the camera is placed at a fixed off-specular angle, some of those irregular reflections due to the presence of defects also will be captured.

The following values were used to generate the following simulation images:

1. a specular fraction ρ_s of 0.9, roughness parameter σ of 0.07;
2. an illumination elevation angle θ_i is fixed at 300 and a view angle θ_r at 0°;
3. defect directions and location are various and set randomly.

For the sake of simple, the camera parameters (e.g., exposure time and illumination intensity, etc.) are not taken into account. For real application, these parameters will be discussed in Sect. 9.5.

Figure 9.16a shows the 3D data of a hemisphere with six defects. These defects are in various location and direction Fig. 9.16b presents two defects section.
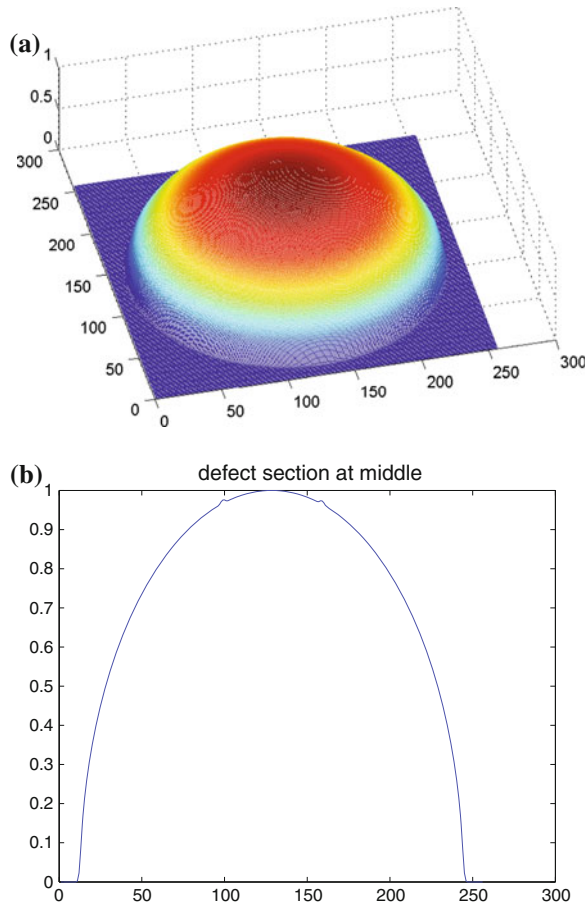
Fig. 9.16 Plots of a hemisphere with six defects **a** 3D plot **b** section plot at middle position

Figure 9.17 shows two set of images from various illumination directions. These simulation images show that: (1) For highly specular reflection surface, the contrast between defect and flawless region significantly depends on defect direction and location. Some defect will appear brightly with intensity of 1, which intensity saturate the CCD. However, some defects appear very low contrast from flawless region. (2) For highly specular reflection curved surface, specular reflection cannot be completely avoided. These bright regions perhaps some cover defect. For this case, any method cannot be used to detect these defects. (3) For highly specular reflection curved surface, all defects with different shape and direction do not appear in anyone image.

Finally, a set of images from real object shown in Fig. 9.18 validate the numerical investigation. The object is a small flat surface and a chrome-plated
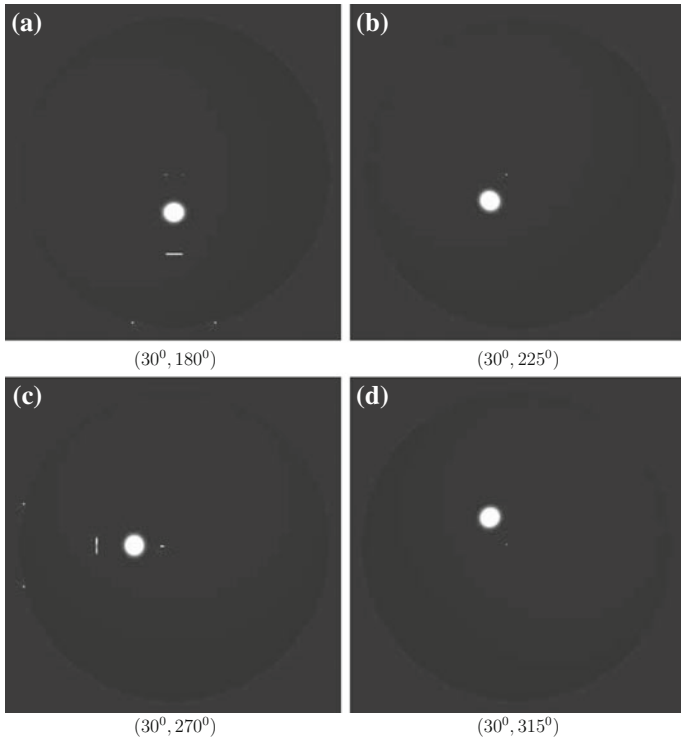
Fig. 9.17 Another set of images from a specular hemisphere with six defects
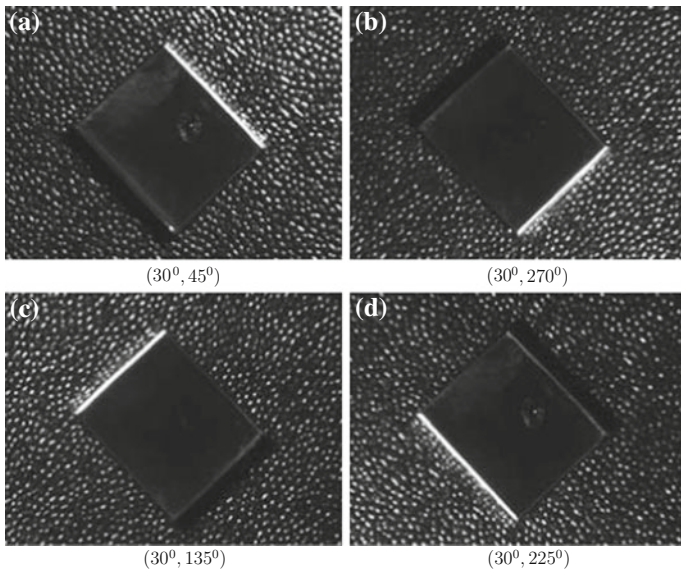
Fig. 9.18 A set of images of a specular flat with a defect

ABS resin surface. (1) Edge in the set of images appears bright in turn. It is obvious that the bright edge is corresponding to the illumination direction. (2) The defect is very bright in some images (e.g., Fig. 9.17a, c) among the set of images. Moreover, the intensity is close to saturation value and high contrast with around region. From these images, defects are easily to be detected. In contrast, the defect is invisible in some images (e.g., Fig. 9.17b, d) among the set of images. It means that it is impossible to extract defects from these images. (3) Whether defects are visible or invisible significantly depends on the illumination when other conditions are fixed. This is validated by numerical investigation and experimental images.

9.3.4 Summary

In this section, a CAD-guided numerical simulation platform has been successfully created. Simulated images can significantly reduce the engineering time to design and successfully implement a vision system. Some numerical simulations are performed on smooth, highly specular flat and curved surface. In addition, a set of images from a real object, which surface is chrome-plate, validate the simulation. (1) It confirms that proper illumination and view direction can help to realize defect inspection on flat specular surface and completely avoid that specular resulted from perfect region enter the view sensor. (2) At the same time, it also confirms that highlight and shadow are unavoidable although the optimum condition is used to capture image. This will make the defect inspection trouble on curved specular surface. However, due to the complex reflection from perfect curved surface or random defects, from a single image it is almost impossible to extract all defects. Another important reason is that the strong specular reflection maybe lead defects covered in highlight regions and the weak diffuse must lead defect appear very low contrast with perfect region. So, in the next chapter, a new novel method is developed to integrate all defects in a single image and eliminate trouble from specular reflection.

9.4 The Specular Reflection Processing on Highly Specular Reflection Surface

Referring to the numerical investigation in Sect. 9.3, because the complex reflection from perfect highly specular reflection curve surface or randomly defects, it is almost impossible to extract all defects from a single image. On the other hand, the specular reflection is very strong while the diffuse reflection components are too weak. In particular, the intensity of the specular reflection often exceeds the limit exposure of the CCD photosensitive surface and results in brightness distortion of

the image. This situation results in that defects information are covered in specular area or defects are not illuminated to produce recognizable defects information. So, in this section, we show a novel multi-image reconstruction method (MIRM) to eliminate the trouble resulted from the specular reflection. Moreover, this method can integrate all defects in a single image. A set of images are captured under various illumination conditions (i.e., different directions) with the fixed object and camera. To the end, a synthetic image is reconstructed in gradient field from the set of images.

9.4.1 Review of Specular Reflection Processing

In computer vision on defects inspection, specular reflection processing has been an active research fields. In general, some researches aimed to separate specular reflection components from the image have been developed. Wolff and Boult [38] used a polarizing filter to separate reflection components from a gray image. The basic idea is that, for most incident angles, diffuse reflection tends to be less polarized than specular reflection. Nayar et al. [39] extended the method above by considering a combination of polarizing filters and colors. A specular pixel has a different intensity if the polarization angle of the filter is changed. Sato and Ikeuchi [40] introduced a four-dimensional space to analyze the diffuse and specular reflection components based on colors and image intensity. This method has the ability to separate the reflection components locally, but it requires dense input images with variation in illuminant directions. Tan and Ikeuchi [41] proposed a specular-to-diffuse mechanism which is based solely on colors, particularly chromaticity, without requiring any geometrical information.

However, the most of them are applied to inhomogeneous objects and there are quite a few methods for processing specular highlights of such perfectly specular surfaces as chrome-plated surfaces. Reflection of metallic surface is different from that of inhomogeneous surface. There are two differences. First, the reflections have different property of polarization. For unpolarized incident light, reflected light from metallic surface is still unpolarized light, while the reflected light from inhomogeneous surface becomes polarized light. Second, for smooth metallic surface, specular reflection is very strong and diffuse reflection is too weak. Moreover, surroundings are often observed as specular reflections. Zhang et al. [20] presented a special random sampling method (SRSM). In this method, a set of images taken in various lighting conditions are converted successfully into synthetic images that contain only diffuse reflection components. Twenty images are taken in different light source directions, so it takes a much longer time to finish the processing. Defect inspection is, respectively, performed in the twenty synthetic images. It is obvious that the specular processing is difficult to apply for real application. On the other hand, in our case, our specular processing is to eliminate the effect from

specular reflection and to inspect defects. So, in this section, a method will be proposed to reconstruct a synthetic image from an illumination sequence images.

9.4.2 The Proposed Specular Reflection Processing

9.4.2.1 The Procedure of the Proposed Method

The procedure of the proposed method is demonstrated as follows. First, a set of images are captured under various conditions (e.g., different lighting directions). Second, the intensity gradient of every input image taken with a certain light source direction is calculated. And then the new gradient image is computed from all input images to eliminate the outliers. Third, a synthetic image is reconstructed from the new gradient image. Last, the resulting image is output for future processing (e.g., defect inspection).

9.4.2.2 The Relation Between Reflection and Lighting Direction

In the Lambertian reflection model (LRM) [42], if both the target object and the viewing position are fixed, the observed image depends only on the lighting direction. The specular lobe component reflected from highly specular surfaces is dominant, as mentioned in the last chapter. Furthermore, the specular lobe component depends mainly on the direction of the light source and the normal direction of the impinged point if all other conditions remain unchanged, as shown in Fig. 9.19. Because the normal direction is vary continuity on a smooth surface, it is obvious that the specular spot reflected from the flawless smooth surface will move smoothly and gradually when a light source moves smoothly.

Fig. 9.19 Specular region shift according to the light direction
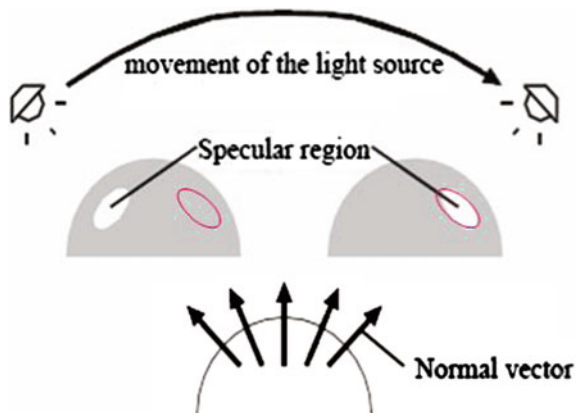
9.4.2.3 Specular Reflection Processing By Multi-image Reconstruction Method

Let $I_k (k \subset [1, n])$ is one of a set of images. I denotes pixel intensity, and sub-index k denotes the sequence number of the image taken with light source direction k. If now a certain pixel (x, y) in all input images on the smooth surface is observed, then the course of the gradient variation of the pixel can be found. If the gradient of the pixel (x, y) is the highest among all input images, the pixel is certainly at a specular boundary. On the other hand, if pixel (x, y) is in a specular region or in a shadow region, then the gradient of the pixel will be close to zero, which is the lowest among all input images. Therefore, the specular pixel can be eliminated by fusing the gradient value of the set of images to get a new gradient image. The new gradient image can be obtained by the following procedure.

1. Calculate the intensity gradient of every input image that is taken with a certain light source direction

$$G_k(x, y) = \nabla I_k(x, y) = \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix}, \qquad (9.18)$$

where k is the image number $k = (1, 2, \ldots, n)$, and ∇ is the gradient operator.

2. Get a new gradient image by weighted summation of the intensity gradient at the corresponding position in all input images, that is:

$$G(x, y) = \sum_k \lambda_k G_k(x, y) = \begin{bmatrix} G_x \\ G_y \end{bmatrix}, \qquad (9.19)$$

where λ_k is the weight of the image k, and Gx, Gy denotes the gradient in directions x and y, respectively.

Some methods of deciding an optimal combination of series gradient images have been adopted based on statistical analysis. Here, a simple and efficient method is used to get the new gradient image. The method is to find the median gradient value at the corresponding position in all of the input images. It is obvious that in the new gradient field, these outliers have been removed and only the gradient values of diffuse reflection components are presented. It needs to be noted that the intensity of a pixel resulting from a defect part of the specular reflection surface, which is similar to a specular component, is due to the sharp change in its normal direction. In addition, these highlights result from a defect present in the same place in most of the series images although the size of a specular spot may differ slightly. It is evident that the processing method can remove specular reflection components because of the intact smooth surface, but results in the preservation of defect information.

One has to reconstruct the synthetic image using the new gradient value obtained in the above process for future processing. One assumes that the synthetic $\tilde{I}$ image satisfies the following equation

$$\nabla \tilde{I} = G(x, y), \tag{9.20}$$

and then the space of all 2D potential functions is searched for a function $\tilde{I}$ whose gradient value is closest to G in the least-squares sense. In other words, the following integral value should be minimized:

$$\int \int F(\tilde{I}, G)\mathrm{d}x\mathrm{d}y, \tag{9.21}$$

where

$$F(\tilde{I}, G) = ||\tilde{I} - G||^2 = \left(\frac{\partial \tilde{I}}{\partial x} - G_x\right)^2 + \left(\frac{\partial \tilde{I}}{\partial y} - G_y\right)^2, \tag{9.22}$$

According to the variation principle, the image $\tilde{I}$ that minimizes the integral value in Eq. (9.21) must satisfy the Euler–Lagrange equation:

$$\frac{\partial F}{\partial \tilde{I}} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial F}{\partial \tilde{I}_x} - \frac{\mathrm{d}}{\mathrm{d}y}\frac{\partial F}{\partial \tilde{I}_y} = 0. \tag{9.23}$$

In fact, Eq. (9.23) is a partial differential equation in $\tilde{I}$.
Substituting F, the following equations obtained:

$$2\left(\frac{\partial^2 I}{\partial x^2} - \frac{\partial G_x}{\partial x}\right) + 2\left(\frac{\partial^2 I}{\partial y^2} - \frac{\partial G_y}{\partial y}\right) = 0. \tag{9.24}$$

Then, the Poisson equation is obtained:

$$\nabla^2 \tilde{I} = \mathrm{div}G, \tag{9.25}$$

where

$$\nabla^2 \tilde{I} = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$
$$\mathrm{div}G = \frac{\partial G_x}{\partial x} + \frac{\partial G_y}{\partial y}, \tag{9.26}$$

is the values divergence of the new gradient.

9.4.3 Image Reconstruction from Gradient Vector Fields

9.4.3.1 Image Reconstruction

For a 2D digital image of size $m \times n$, the multi-grid method can be used to solve Eq. (9.25) and get the synthetic image. In order to satisfy the requirement of the method, the images have to be padded to make them square and of size N being power of two. Lastly, the resulting image is cropped back to its original size.

The image can be regard as a 2D region Ω as shown in Fig. 9.20, where each pixel is regarded as a grid node with an interval distance of h. Let $\tilde{I}(i,j)$ be the intensity of the pixel (i,j) in the reconstructed image $\tilde{I}$. By using the two-dimensional Taylor function, $\tilde{I}(i,j)$ can be represented as the following equation using the neighboring four pixels' intensities:

$$\tilde{I}_{i-1,j} = \tilde{I}(i-h,j)$$
$$= \tilde{I}_{i,j} - h\left(\frac{\partial \tilde{I}}{\partial x}\right)_{i,j} + \frac{h^2}{2}\left(\frac{\partial^2 \tilde{I}}{\partial x^2}\right)_{i,j} - O(h^3), \quad (9.27)$$

$$\tilde{I}_{i+1,j} = \tilde{I}(i+h,j)$$
$$= \tilde{I}_{i,j} + h\left(\frac{\partial \tilde{I}}{\partial x}\right)_{i,j} + \frac{h^2}{2}\left(\frac{\partial^2 \tilde{I}}{\partial x^2}\right)_{i,j} - O(h^3), \quad (9.28)$$

Fig. 9.20 Two-dimensional grid
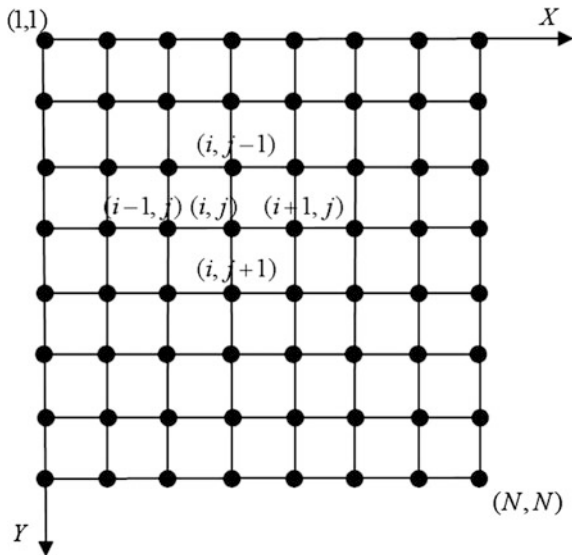
where $O(h^3)$ is the remainder term of third order. By adding Eqs. (9.27) and (9.28), we can obtain the second order $\frac{\partial^2 \tilde{I}}{\partial x^2}$ in the x direction.

$$\left(\frac{\partial^2 \tilde{I}}{\partial x^2}\right)_{i,j} = \frac{1}{h^2}\left(\tilde{I}_{i-1,j} - 2\tilde{I}_{i,j} + \tilde{I}_{i+1,j},\right. \tag{9.29}$$

In the same way, the second-order partial derivatives $\frac{\partial^2 \tilde{I}}{\partial y^2}$ can be obtained in the direction:

$$\left(\frac{\partial^2 \tilde{I}}{\partial y^2}\right)_{i,j} = \frac{1}{h^2}\left(\tilde{I}_{i,j-1} - 2\tilde{I}_{i,j} + \tilde{I}_{i,j+1}\right). \tag{9.30}$$

Let h be the unit distance ($h = 1$), and adding Eqs. (9.29) and (9.30), the following equations are obtained:

$$\begin{aligned} \nabla^2 \tilde{I} &= \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \\ &= \tilde{I}_{i-1,j} + \tilde{I}_{i+1,j} + \tilde{I}_{i,j-1} + \tilde{I}_{i,j+1} - 4\tilde{I}_{i,j}. \end{aligned} \tag{9.31}$$

Using a similar method, one can also represent div$G$ as in the following equation:

$$\begin{aligned} \mathrm{div}G &= \frac{\partial G_x}{\partial x} + \frac{\partial G_y}{\partial y} \\ &= G_{x_{i,j}} + G_{x_{i-1,j}} + G_{y_{i,j-1}} + G_{y_{i,j+1}}. \end{aligned} \tag{9.32}$$

In order to solve Eq. (9.25), the reconstructed image $\tilde{I}$ has to be converted to a matrix U in the direction y or x.

$$U = [\mathbf{u}_1, \mathbf{u}_1, \ldots, \mathbf{u}_{N-1}]^T, \tag{9.33}$$

where $\mathbf{u}_j = [\tilde{I}_{1,j}, \tilde{I}_{2,j}, \ldots, \tilde{I}_{N-1,j}], j = (1, 2, \ldots, N-1)$.

At the same way, one can also convert div$G$ to a matrix F in the direction y.

$$F = [\mathbf{f}_1, \mathbf{f}_1, \ldots, \mathbf{f}_{N-1}]^T, \tag{9.34}$$

where $\mathbf{f}_j = \frac{1}{4}[\mathrm{div}G_{1,j}, \mathrm{div}G_{2,j}, \ldots, \mathrm{div}G_{N-1,j}], j = (1, 2, \ldots, N-1)$.

Then, the linear equation shown in Eq. (9.35) can be obtained by substituting Eqs. (9.30)–(9.34) into Eq. (9.25).

$$A \times U = F, \tag{9.35}$$

where A is an order $(N-1)^2 \times (N-1)^2$ sparse matrix and can be shown as follows:

$$A = \begin{bmatrix} K & -E/4 & \cdots & 0 \\ -E/4 & K & -E/4 & \vdots \\ \vdots & \ddots & \ddots & -E/4 \\ 0 & 0 & -E/4 & K \end{bmatrix}, \qquad (9.36)$$

where E is a $(N-1)^2$ order unity matrix, and K is a square matrix of order $(N-1)^2$ and can be shown as follows:

$$K = \begin{bmatrix} 1 & -1/4 & \cdots & 0 \\ -1/4 & 1 & -1/4 & \vdots \\ \vdots & \ddots & \ddots & -1/4 \\ 0 & 0 & -1/4 & 1 \end{bmatrix}. \qquad (9.37)$$

For this large-scale linear system, the convergence speeds of the Jacobi and Gauss–Seidel iteration methods are very slow. The multi-grid method is an efficient algorithm for accelerating convergence of iterative methods. The scheme alternates between an iteration that quickly reduces local (high frequency) errors, and an approximate solution on a coarse grid for reducing the global (low frequency) error.

9.4.3.2 The Multi-grid Method Based on FDM

The iteration procedure is conducted in the pyramid-like structure as shown in Fig. 9.21, in which the lowest level is the most refined grid $N \times N$, and grids in higher levels are successively coarser grids. The method is abbreviated as 'FDM(MG).'

Fig. 9.21 Scheme of multi-grid algorithm
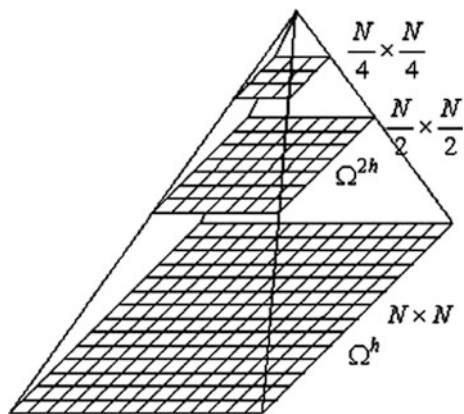
The basic procedure of the multi-grid method is summarized in the following algorithm [43]:

Step 1   On the finest grid space Ω^h, which is the lowest level shown in Fig. 9.21 and whose grid interval is h (it can be defined as unity), one relax γ_1 times to solve Eq. (9.35) with an initial guess $\tilde{U}_0^h$ and get the approximate $\tilde{U}^h$.

Step 2   On the coarser grid space Ω^{2h}, which is the second level from the bottom shown in Fig. 9.21 and whose grid interval is $2h$, the residual Eq. (9.38) is solved and the error value v^h is obtained. I_h^{2h} is defined as a restriction operator for mapping the residual value from fine grid space Ω^h to the coarser grid space Ω^{2h}.

$$L^{2h} \times v^h = r^{2h}, \tag{9.38}$$

where $r^{2h} = I_h^{2h}(F^h - L^h \times U^h)$ and $L^{2h} = I_h^{2h} L^h I_{2h}^h$.

Step 3   In this step, the following Eq. (9.39) is used to correct the fine grid solution $\tilde{U}^h$ by the error value v^{2h}:

$$\tilde{U}^h = \tilde{U}^h + I_h^{2h} \times v^{2h}. \tag{9.39}$$

Step 4   Backing up to the finest grid space Ω^h, relax γ^2 times to solve Eq. (9.35) using the approximate $\tilde{U}^h$ as initial solution and get the solution.

The other methods, such as discrete wavelet transform (DWT) [44] and discrete Fourier transform (DFT), also can be used to solve a discretization of 2D Poisson Eq. (9.25), but they are omitted in this section (Fig. 9.22).

Fig. 9.22   Original image of the Lena

9.4.3.3 Evaluation of the Multi-grid Method

This section presents confirm test results of the FDM(MG). The algorithm described above was implemented with one synthetic image and a real image. The Gaussian noise (with a mean set zero and a standard deviation set to 0.01) was subsequently added to the images. Next, above the multi-grid method is evaluated by using of the Lena image. Figure 9.23 shows the original image of the Lena. Figure 9.24 shows the reconstructed image by FDM(MG). In addition, one provides quantitative measures on how well the reconstructed image matches the original by looking at the mean square error (MSE) and the relative MSE. The MSE and relative MSE for the above approaches are shown in Table 9.2.

9.4.4 The Application on Highly Specular Reflection Surface

9.4.4.1 Specular Processing Results Obtained from the Model

First, a simulation image is generated from a vase, which explicit surface equation is presented by the Eq. (9.7): Four synthetic images are generated by the reflection model and its surface shape 3D data shown in Fig. 9.9. The four images are shown in Fig. 9.24, where the illumination elevation angle was fixed at 30° and their azimuth angles were 30°, 120°, 210°, 300°, respectively.

Figure 9.25 shows the reconstructed image and Fig. 9.26 shows the histogram comparison between the original image with the specular processing image. However, left image of Fig. 9.26 shows the histogram of the original image shown in Fig. 9.24a, and right image of Fig. 9.26 shows the histogram of the synthetic

Fig. 9.23 Reconstructed image by FDM(MG)

(a)   **(b)**

$(30^0, 30^0)$ $(30^0, 120^0)$

(c)   **(d)**

$(30^0, 210^0)$ $(30^0, 300^0)$
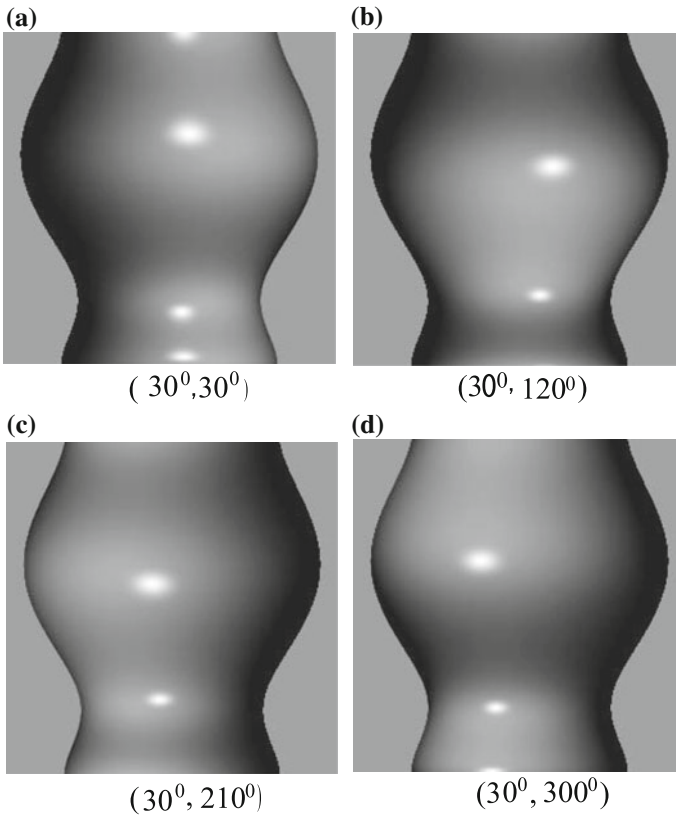
Fig. 9.24 Four images of a vase generated by its 3D data and reflection model

image shown in Fig. 9.25. It can be seen from left image of Fig. 9.26 that many highlight reflections can be observed while the synthetic image shown in Fig. 9.25 shows there are few or no pixels with high-intensity values because the specular reflection components are significantly removed by the above-proposed method. On the other hand, the histogram of the original image is a non-normal distribution which shows the brightness distribution is partially bright, dark, or of excessively concentrated brightness. The histogram of the synthetic image follows the normal distribution, which indicates the synthetic image is favorable to further processing.

Table 9.2 Comparison between various reconstruction methods

Methods	MSE	RMSE	Time (s)
FDM	9.378	38.17	0.232
DFT	0.003	0.031	0.405
DWT	0.302	3.636	18.08
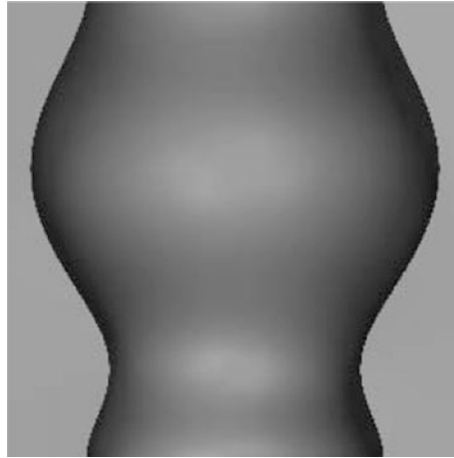
Fig. 9.25 The vase's image after specular processing
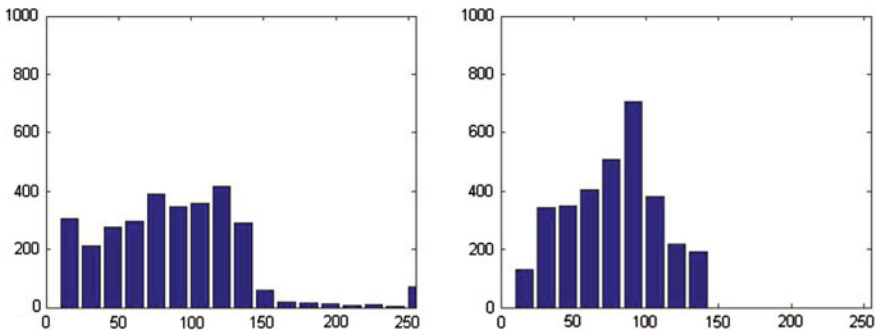
Fig. 9.26 Histograms comparison between the original and synthetic images

9.4.4.2 Specular Processing Results Obtained from Real HSR Curved Surface

Figure 9.27 shows four original gray images of the ABS resin, which were taken with different incidence light directions. The synthetic image is given in Fig. 9.28, which was reconstructed using the proposed method. The histogram of the first image shown in Fig. 9.27a is given in left image of Fig. 9.29. In each original image shown in Fig. 9.27, some pixels are very dark, whose intensities are close to 0, while some pixels are very bright, which are close to or equal to 255 in intensity. This is because, for chrome-plated surfaces, diffuse reflections are very weak and also the specular reflections are very strong. However, it is obvious that entire surface of the synthetic image appears with medium gray level. The histogram of the synthetic image is presented in right of Fig. 9.29. Comparing the two histograms

$(30^0, 45^0)$ $(30^0, 135^0)$

$(30^0, 225^0)$ $(30^0, 315^0)$
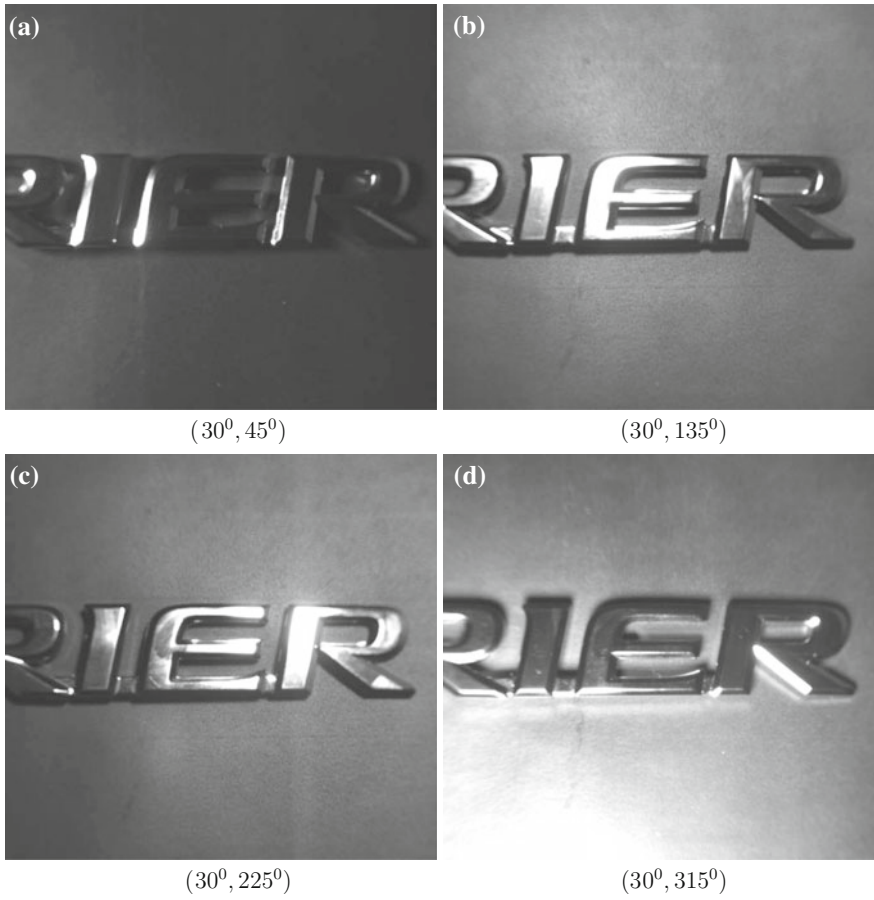
Fig. 9.27 Four original images of the ABS resin taken with different light direction

Fig. 9.28 The synthetic image of the ABS resin
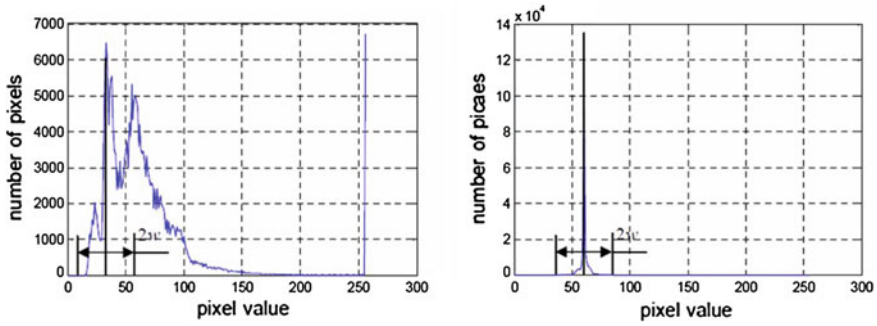
Fig. 9.29 Histograms comparison between the original and synthetic images of the ABS resin

of left and right of Fig. 9.29, it can be observed that the high-intensity bright spots have been completely removed from the original image, which has the corresponding histogram spike at pixel value 255.

To compare the intensity distribution between the original image and the image after specular processing, we first set a specific area which centers around the intensity line for peak point and widths $2w$. Then, we calculate the intensity percentage in the condition of the different band width $2w$ and show results in Fig. 9.30. However, the square marks show the width $2w$ obtained from the histogram of the original image, and the triangle marks show the width w obtained from the histogram of the synthetic image. It can be seen from Fig. 9.30 that after the intensity in the special area rises to more than 0.95, the synthetic image has a better distribution than that of the original image. In the other words, in a wide area, the intensity distribution for diffuse reflection components becomes worse when removing specular reflection components. This means that the clarity of the synthetic image is better than that of the original image. In addition, the synthetic

Fig. 9.30 Intensity distribution between the original and the synthetic image of the ABS resin
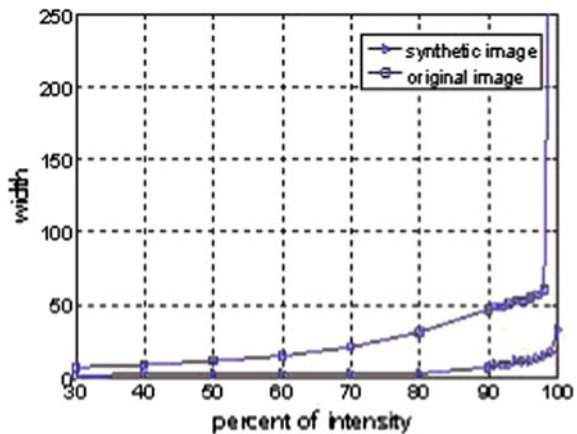
image histogram has a reduced width as a result of a reduction in the higher pixel values in the main histogram peak.

9.4.5 Summary

Referring to the numerical investigation in Sect. 9.3, because the complex reflection from perfect highly specular reflection curve surface or randomly defects, it is almost impossible to extract all defects from a single image. So, this section has proposed a new approach to integrate all defects in a synthetic image and eliminated negative percent w of original image w of synthetic image effect from specular reflection. A set of images of the object are captured with different lighting directions and then reconstruct a synthetic image from the new gradient value calculated from the set of images. The approach of multi-grid method based on FDM was especially discussed to reconstruct the synthetic image. By confirm tests, the following conclusions can be obtained.

1. The specular spot reflected from the flawless smooth surface will move smoothly and gradually from various images.
   The specular lobe component reflected from highly specular surfaces is dominant. Specular lobe component depends mainly on the direction of the light source and the normal direction of the impinged point if all other conditions remain unchanged. So, the specular spot reflected from the flawless smooth surface will move smoothly and gradually from a set of images image. The set of images are captured under significant different illumination direction while others are fixed. This observation is the base of the MIRM approach proposed in this chapter.
2. The synthetic image shows some positive changes comparing with the original images.
   The synthetic image appears more uniform gray change because that the specular regions with saturation intensity are removed. This change is undeniably favorable to extracting defects from the synthetic image. But, in original images, some pixels appear too dark or too bright. All defects have been successfully integrated in the synthetic image and all defects are integrated in a single image. It is that all defects can be extracted from the synthetic image. The specular processing method proposed in this chapter can be used as the pro-processing of defect inspection.
3. The method proposed in this section can be used as the pro-processing of defect inspection.
   The image reconstruction method is robust and short time-consuming, and it is feasible to use as pro-processing of defect inspection in real application. So, in the next chapter, a new vision defect inspection system will be presented.

9.5 A New Defect Inspection Method

A synthetic image has been reconstructed from a set of images by the method proposed in the last section. All defects have been successfully integrated in the synthetic image. So, this section focuses on how to extract defects from the synthetic image. Although the specular regions with saturation intensity are successfully removed, the synthetic image includes kinds of noise and some areas with uneven intensity. This uneven intensity would perturb image segmentation. So, morphology operation is to eliminate uneven pixels. On the other hand, it is very difficult to extract object silhouette from background due to the complexity of object edge and too weak diffuse reflection of the object, as well as traditional template matching technique based on gray image is sensitive to the relative position of the inspection part and the template.

So, a new defect inspection method is proposed to extract defects from the synthetic image. In this defects extraction algorithm, the template matching is performed binary images of the template and the inspection part. Moreover, the morphology technique is adapted to eliminate uneven pixels and decrease the influence of the relative position gap between the template and the inspection part. In addition, the defect inspection procedure was introduced in detail by a complete defect inspection experiment. It confirms that the defect inspection method is feasible on highly specular reflection surface. In particular, the method is low algorithm complexity and is able to meet the real-time defect inspection.

9.5.1 Review of the Traditional Defect Extracting Methods

All defects have been integrated in a single image obtained after specular processing, and the trouble resulted from specular reflection also has been successfully eliminated. So, the main objective of this section is to extract defects from the synthetic image. Defects are viewed as in-homogeneities in regularity and orientation fields. Two distinct but conceptually related approaches are presented. The first one defines structural defects as regions of abruptly falling regularity, and the second one as perturbations in the dominant orientation. Both methods are general in the senses that each of them is applicable to a variety of patterns and defects [45]. Referring to above principle, many papers [46–50] have been investigated the method of defect extracting and classification. Liu et al. [51] proposed a spectral subtraction method to successfully applied to image processing and to detect defects in the integrated circuit (IC) image. Based on morphology, Reference [52] presents the method called Morphological Filtering for defect detection. I this section, we first calculate the difference between the IC image with defects and the standard image without defects. Then, we digitize the differential image to two gray levels. The binary image is processed by means of mathematical morphology to detect defects. This method requires the standard image to be presented in advance.

Reference [53] proposes the method called a quantile–quantile plot based on the pattern matching to detect defects. A pattern-matching image based on the quantile–quantile plot (Q–Q plot) is proposed, and the method also needs the standard image.

9.5.2 The Proposed Defect Extracting Method

The conventional template matching technique is to subtract the sample image from the template image according to gray level. It is sensitive to noise and misalignment between the two images. On the other hand, in our application, a little specular reflection in such locations as edges and cross points also affects the result from the traditional approach method. So, based on conventional a technique [52], a new defect inspection approach is proposed to lead efficient defect segmentation in this section. This approach proposed combines template matching and morphology technique to overcome the shortcomings of the traditional method. In addition, its low computation complexity is a considerable favorable property to real-time defect inspection.

9.5.2.1 Review of Morphology Operations

The morphological operation is one of a number of operations consisting of combinations of the Minkowski sum and Minkowski difference of two sets A and B, which can be shown in Eq. (9.40), respectively. In the case of image processing, the set A is called the target element and set B is the structural element

$$
\begin{aligned}
A \oplus B &= \{z \in E : z = a + b, a \in A, b \in B\}, \\
A \ominus B &= \{z \in E : z - b \in A, b^{\forall} \in B\}.
\end{aligned}
\tag{9.40}
$$

Generally, the Minkowski sum calculation is called Dilation and the Minkowski difference calculation is called Erosion. Furthermore, performing Dilation after Erosion is called opening and performing Dilation before Erosion is called closing. Figure 9.31 shows an example of the morphological operations in the case of a binary image. As shown in Fig. 9.31, Dilation causes the target figure to expand and Erosion causes the target figure to shrink. Opening removes only the convexities and isolated regions of the target figure, and closing removes only concavities and holes. Assume a grayscale image has horizontal axis x, vertical axis y, and let the brightness be represented by $f(x, y)$. In one dimension, such as along the axis x, the Dilation and Erosion operations are shown as Eq. (9.41).

$$
\begin{aligned}
(f \oplus g)(x) &= \max_{x-z \in F, z \in K} \{f(x - z) + k(z)\}, \\
(f \ominus g)(x) &= \min_{x-z \in F, z \in K} \{f(x + z) - k(z)\},
\end{aligned}
\tag{9.41}
$$
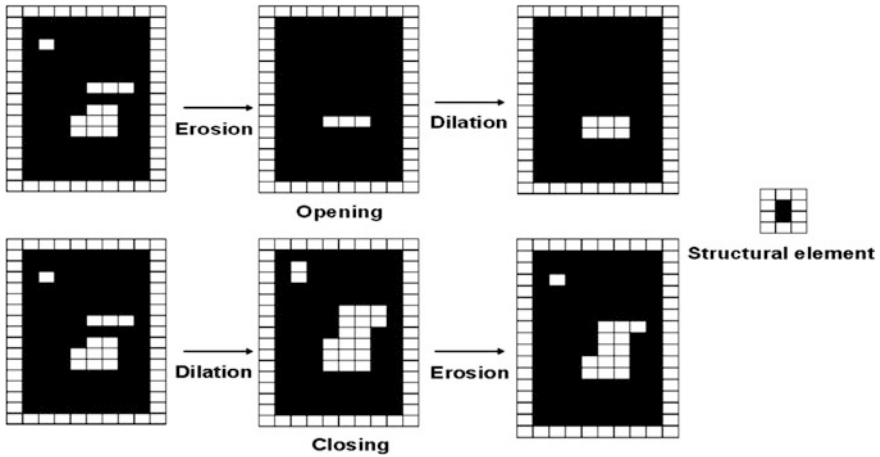
Fig. 9.31 Example of the morphological operation applied to a binary image

where $f(x)$ denotes the original image, $k(z)$ is the structural element, and F and K show the domain of the structural element.

Figure 9.32 shows an example of the morphological operation for a grayscale image, where figure (a) shows $f(x)$, (b) shows the structural element $K(z)$, (c) shows the result obtained from Eq. (9.41) and (d) shows the result obtained from Eq. (9.41). As shown in Fig. 9.32, in the case of the Dilation operation, pixels whose intensity values are much higher than those of neighboring pixels and whose area is smaller than the structural element are removed. In contrast, in the case of

Fig. 9.32 Example of the morphological operation for a *grayscale* image **a** f(x) **b** K(z) **c** dilation **d** erosion
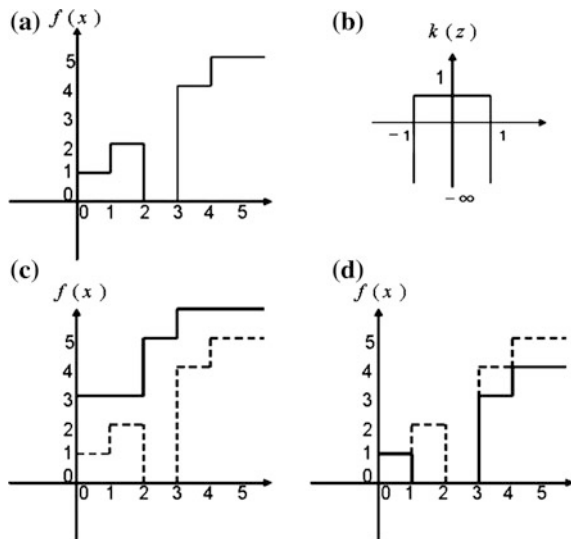
the Erosion operation, pixels whose values are much lower than those of neigh-
boring values and whose area is smaller than the structural element are removed.

9.5.2.2 Defect Detection Algorithms by Combining Morphology Operation and Template Matching

Figure 9.33 shows the defect inspection procedure proposed. In the first step, the
opening operation, difference, and image segment operation are performed. And
then, the image obtained in the above operation is converted to a binary image by
using a suitable threshold. The binary image of the investigation object is set as
bwI_s. In the same way, the template image is constructed from the flawless surface
and it is set as bwI_t. Lastly, all defects are detected by subtracting the template
binary image bwI_t from the sample binary image bwI_s according to Eq. (9.42).

$$D(i,j) = bwI_s(i,j) - bwI_t(i,j), \tag{9.42}$$

where $D(i,j)$ represents the difference between bwI_s and bwI_t. Consequently, any
pixel whose value is one as being in a defect. Some holes have to be filled in the
resultant binary image D to make a close region and eliminate the noise. On the
other hand, those regions whose areas are larger than a set threshold are classified
into defects.

In addition, one need to dilate the template binary image before performing the
subtract calculation. This operation can guarantee that the edge of the template is
wider than in the sample image. Moreover, when there is a small difference in
location between the sample and template, the approach is also robust. That is, in
this way, the drawback of template matching being sensitive to changes in the
relative position of the inspection part and the template, and being sensitive to
changes in the photographic environment can be overcome.

9.5.2.3 A Complete Inspection Experiment

A complete inspection on a highly specular reflection (e.g., Chrome-plated) curve
surface is introduced to describe the inspection procedure in detail in this
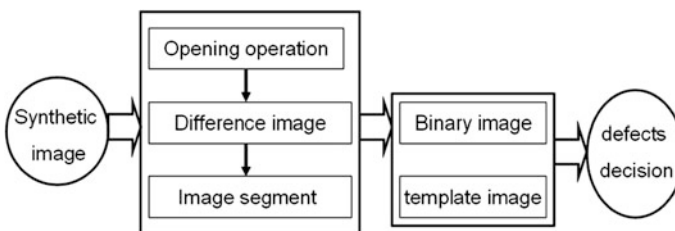
Fig. 9.33 The defect inspection procedure from the synthetic image

subsection. On the surface, two defects are detected by human operator. For the sake of the next illustration, the defect marked by the circle is called defect *A* and the defect marked by the rectangle is defect *B*. As shown in Fig. 9.34, defects *A* and *B* are of different shapes, where *A* is a small dot and *B* is a thin scratch. And the resolution of the inspection system is set at 0.075 mm/pixel according to the inspection standard and morphology technique. The object is illuminated by four different directions in turn, and the illumination elevation angle θ_i is set at 30° and the azimuth angles φ_i are set at (a) 30°, (b) 120°, (c) 210°, and (d) 300°, respectively.

Figure 9.34 shows four original gray images of size 640 × 480 pixel captured from the chrome-plated surface. From them, the following case can be observed. First, in each original image shown in Fig. 9.34, most of the pixels are very dark, whose intensities are close to 0. It can be confirmed that the illumination system successfully avoids specular reflection from a flawless surface. So, it can indicate that the lighting system with side and diffuse illumination succeeds in reducing the specular reflection from a smooth surface, although some brightness appears at the edge. Second, it is obvious that their reflections strongly depend on the illumination direction, so all defects cannot be detected from any one image. From the above discussion, one can say that all defects could be detected only from multiple images captured with different directions. Third, specular area completely move according to the illumination direction vary. This case can ensure the effect from specular area can be eliminated.
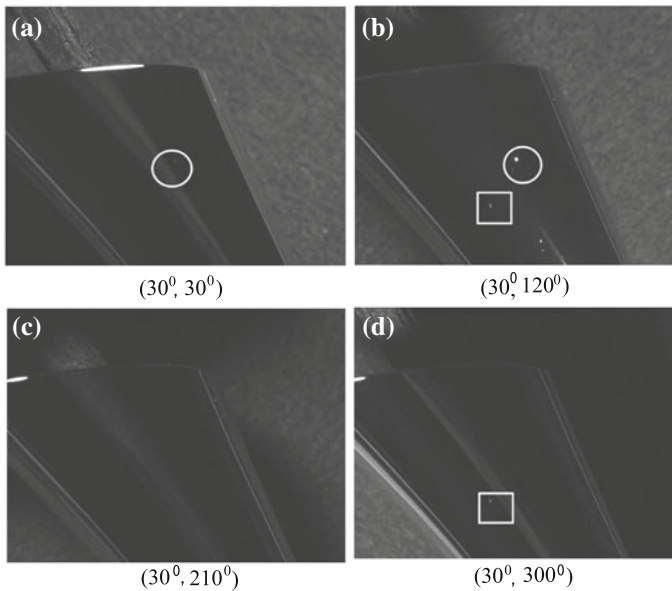
Fig. 9.34 Four original images of chrome-plated surface
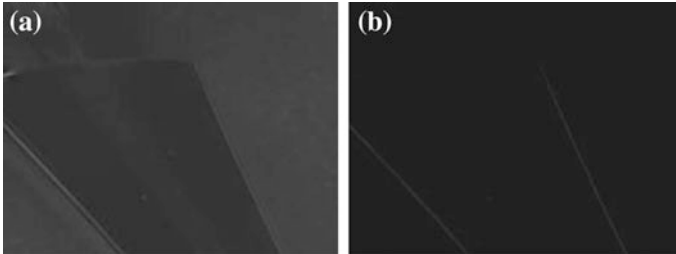
Fig. 9.35 The synthetic image and morphology processing result. **a** The synthetic image of the object. **b** The morphology processing image

Figures 9.35, 9.36, and 9.37 show the processed results by the proposed method. To evaluate the specular processing, the synthetic image is compared with the original images of Fig. 9.34. For example, as shown in Fig. 9.34a, only defect A appears with very low contrast, and the worst case is that defect B does not appear. And a highlight band appears on the top edge. In particular, all defects cannot be found in Fig. 9.34c. In contrast, in the synthetic image, the highlight band was completely removed and the two defects simultaneously appear with considerably high contrast. On the other hand, in the center of the synthetic image some uneven intensity can be found, which will make the next defect inspection difficult or erroneous. So, the morphology operation is adopted to eliminate this phenomenon. Figure 9.35 shows the morphology operation result and the improvement can be confirmed. Figure 9.36 shows defect detection results by the proposed method, where (a) shows the binary image of the sample part and (b) shows the binary image of the template part. Furthermore, Fig. 9.37a shows the result of template matching obtained by subtracting the target image from the template image, and Fig. 9.37b shows the final result after noise removal. In order to guarantee the effectiveness of the proposed method, we have to dilate the template image to enlarge the edge size. This means that when the target and the sample are slightly out of alignment, our method is still effective. Furthermore, the car manufacturer set a manufacturing standard such that objects with defect sizes of more than 0.25 mm are to be rejected. According to the standard and system resolution of 0.075 mm/pixel, those small white spot noises are removed by the erosion
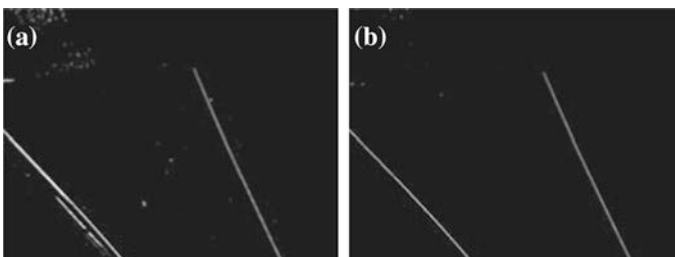
Fig. 9.36 Binary image of the object and the template. **a** The binary image of the object. **b** The binary image of the template
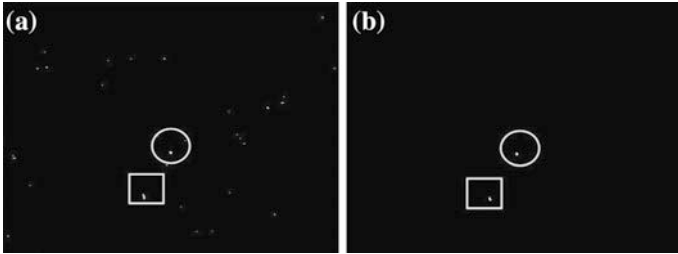
Fig. 9.37 Defects segmentation and recognition. **a** The result of template matching. **b** All defects on the object

algorithm using a square structure element. To the end, two defects are successfully detected.

9.5.3 Summary

A new defect inspection method is proposed to achieve the defect inspection in this section. All defects can be efficiently extracted from the synthetic image in which all defects are gathered. In this defects segmentation algorithm, the template matching is performed on the binary image of the template and the inspection part. Moreover, the morphology technique is adapted to eliminate uneven pixels and decrease the influence of the relative position gap between the template and the inspection part. So, the method is able of overcoming the shortcomings of the traditional method. In addition, a complete defect inspection experiment is presented to show the defect extracting procedure in detail. It confirms that the defect extracting method is feasible and robust. In particular, the method is low algorithm complexity and is able to meet the real-time defect inspection.

9.6 A Vision Defect Inspection System and Its Application on HSR Curved Surface

Based on the previous section analysis on highly reflective surface (HSR), in this section a new vision defect inspection system and its application on HSR curved surface is developed. The inspection system is comprised of the following components: (1) an acquisition system; (2) an inspection platform; and (3) a host PC. The system is capable of revealing, detecting, and characterizing on curved surface with strong specular reflection while very weak diffuse reflection. In addition, two typical examples of kinds of experiments are presented and discussed. In the first experiment, the object called cover-knob is chrome-plated surface [20], which acts like a mirror. The object pose and illumination parameters are analyzed and

determined by means of the simulation platform developed in Sect. 9.3. In addition, various experiments different illumination conditions are performed to evaluate the acquisition system by analyzing the image quality. All defects are robustly detected in lots of experiments. Experiments confirm that better image with less specular reflection from flawless region and high defect expressivity is the very important and efficient to defect inspection. In the second experiment, the object is car doorknob. Specular reflection from the surface is big area and cannot be efficiently avoided, so two group images are captured to reconstruct an image in which all defects are successfully integrated. Due to the more uniform intensity and high defect expressivity, all defects also are successfully detected.

9.6.1 The New Vision Defect Inspection System

To accomplish the automated inspection, a new vision defect inspection system has been developed. The system is capable of revealing, detecting, and characterizing on curve surface with strong specular reflection while very weak diffuse reflection. Figure 9.38 shows the schematic representation of this proposed system, which consists of the following components:

1. An image acquisition system, which includes an camera generally fixed on the above of inspection platform and an illumination system that includes a group of independent light source. The camera and its setting and the light source and its parameters of the image acquisition system will be further discussed in detail.
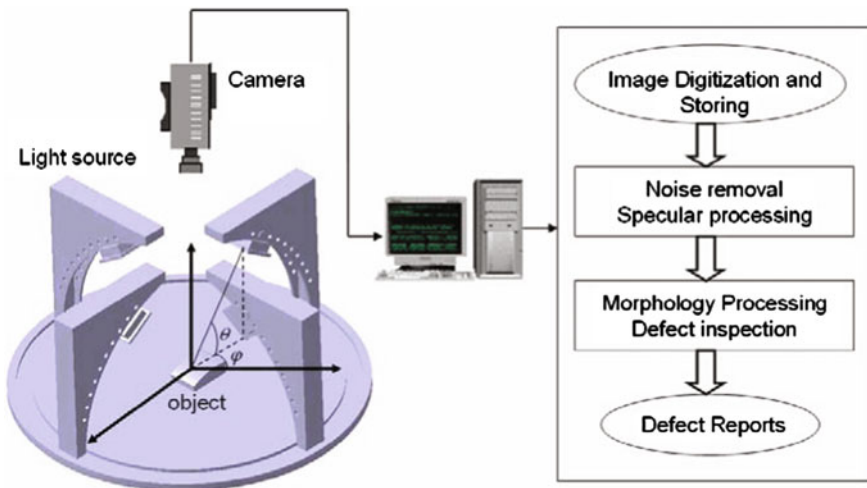
Fig. 9.38 The schematic representation of the automatic vision inspection system

2. An inspection platform, on which inspection object is fixed. In addition, the object pose can be adjusted to satisfy inspection requirement or realize all surface inspection of the part.
3. A host PC to mainly compute the object pose, illumination parameters and to process the images and output inspection result.

The prototype machine of the inspection system proposed is shown in Fig. 9.39.

9.6.1.1   Camera and Its Setting

For the camera the most popular sensor, the charge coupled device is adopted. The camera is Toshiba Teli CSFV90BC3 [54]. It is popular among industrial users mainly because of its high reliability, shock resistance, high sensitivity, resistance to electromagnetic interference, and the linearity of its response with respect to luminance. The CCD camera which provides an 8-bit gray level image was mounted in a fixed position perpendicularly above the work piece. An admissible viewpoint and CCD should satisfy the following constraints that are checked in the following order:

1. Visibility: This constraint ensures that there is no occlusion between the viewpoint and the lens surface to be inspected;
2. Field of view: Vision sensors have limited field of view by the size of the sensor area and the focal length of the lens. The homogeneous transformation matrix is used to compute the surface points that will be projected inside the sensor area.

Fig. 9.39 The prototype machine of the inspection system proposed
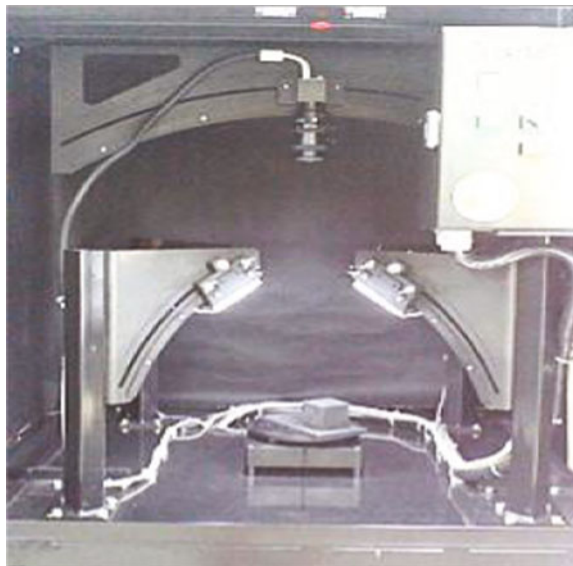
This mathematical modeling gives a direct relation between the camera and world coordinate systems. So, the image coordinates of a point x in the world coordinate system is obtained from: $x = TX$, where T is homogeneous transformation matrix, and it relates the vision sensor and the world coordinate systems through a rotation, translation, and perspective transform.

3. Viewing angle: It implies that the curvature of the field of view can be chosen by selecting a limit between the viewing direction and the normal of the surface points;

4. Resolution: Ensures that the smaller size of defect accepted on every inspection surface area with length l is imaged to at least p pixels on the image plane. Assume that the vision sensor is posed normal to the lens surface, the number of active pixels 1 s corresponding to a defect is a linear function of the defect occupancy surface. Hence, using the pinhole lens model of the camera shown in Fig. 9.40, a theoretical link, pixel to feature size, can be computed as Eq. (9.43):

$$x_s = \frac{d}{f} p x_c, \qquad (9.43)$$

where x_s is the scene feature size, d is the distance from the vision sensor to the lens surface, f is the focal length, p is the number of active pixels, and x_c is the pixel size.

9.6.1.2 The Light Source System

In computer vision applications, the provision of a correct and high-quality illumination is absolutely decisive [55]. Hence, the light needs to be provided in a control manner to accentuate the desired features of the surface avoiding disturbances which could alter the quality of the acquired image. So, an inadequate

Fig. 9.40 Pinhole model to set resolution

$$\overline{O_c O} = f$$

$$\overline{O_s O} = d$$

$$\overline{O_c C} = x_c$$

$$\overline{O_s S} = x_s$$
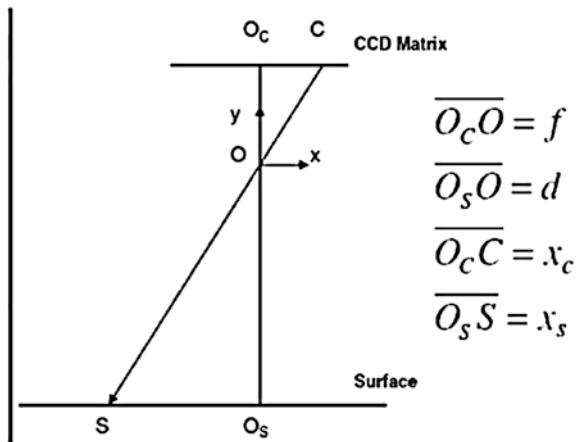
lighting system selection involves the development of complex computer vision algorithms to extract information from the images, or even imply an unfeasible vision task [56]. There is a rich variety of lighting techniques that may be used in machine vision. From the available commercial lighting systems, the most recommended for this application are the backlight and the diffuse dark field systems and structured lighting.

Light diffusion produces an area of uniform illumination that helps to eliminate shadows and minimize specular reflection. Diffusing light is achieved by placing a Mylar screen in front of a light source or by reflecting the light back onto a surface using a white screen, thereby enlarging the area of coverage by the light source. Such diffusion assemblies can use a number of different lighting techniques, such as fiber, fluorescent, or light-emitting-diode illumination.

In our system, a square LED with model of PDF-7X7CWKC has been used. It is produced by Japan Kondoh Seisakusho Co., Ltd. The light emitted from LED passes through frosted glass and provides intense uniform light in entire direction. So, it can be approximately considered as diffuse light source. In this system, a simple and efficient illumination system has been developed to detect defects on strong reflective surface. It aims at enhancing defects on the inspection surface and avoid specular area appear in the image acquisition. Since the specular reflections of perfect specular reflection surface are very strong and diffuse reflections are weak, high-quality image cannot be captured by means of conventional illumination light source. In the proposed illumination system, square LED with white light and a separate constant current source with intensity control were adopted. The light emitted from LED passes through frosted glass and provides intense uniform light in entire direction. So, it can be approximately considered as diffuse light source. The light hit the surface to be detected under an optimum location to avoid specular reflection from the direct illumination method. Moreover, the illumination system consists of several light sources, which are mounted on a rack according to the optimum location.

On the other hand, diffuse illumination source can eliminate shadows and greatly reduces the effect of the specular reflections. Moreover, diffuse illumination source with a concern direction illuminate specular surface also can reduce specular reflection from the surface and enhance the edge brightness. Many experiments on the specular reflection surfaces, such as the chrome-plated surfaces, show that the defects detected are clear compared with surrounding region under proper light intensity and incidence direction. However, highly specular surface with complicate shape is highly reflective, and specular reflection can be not completely avoided even by using a uniform diffuse light source. On the other hand, variety of defects which have different geometry shape may be illuminated to perform specular reflection or covered by bright region. These will affect severely the detection of the defect detection system accuracy. The side illumination can avoid specular reflection from the direct illumination method. It is favorable for inspecting objects with highly reflective or uneven surfaces. Based on the above discussion, one confirm that all defects on HSR curve surface cannot be obtained correctly and completely

from a single image. So, as shown in Fig. 9.38, an illumination system has been designed specially in the proposed inspection system.

A group of light sources are mounted on their rack, respectively. The light elevation angle is very important to obtain high-quality image. The elevation angle can be adapted according to the surface to be objected. Generally, the group of light sources should be even apart and of fixed slant. The optimized parameters of light source which mainly depend on the shape and optical properties of the object to detected play an important role to obtained favorable images. The elevation angle of light direction is determined by the CAD-guide platform proposed in Sect. 9.3 by combining reflection properties of surface to be detected, light source parameters, and object position. In the real application, the light source for the investigation object has to be set especially due to the complexity of surface shape and the random of defects.

9.6.2 Image Quality Evaluation

In computer vision defect inspection, the provision of a high-quality image is absolutely decisive. In particular for inspection on highly specular surface, specular reflection often saturate the CCD to lead surface information loss or hide defects feature. So, in this section, the area ratio of saturation pixel is proposed to assess the image quality. On the other hand, it is also very important to capture high contrast between defects and perfect region. So, the defect expressivity is proposed to assess the image quality. By comparing the area ratio and defect expressivity, the optimum image acquisition condition can be determined. Diffuse light source can illuminate objects uniformly. But for highly specular reflection curve surface (e.g., chrome-plated surface), it is very important to set illumination location to guarantee less specular reflection entering the camera and high-quality image favorable to defect inspection. It mainly depends on the macro-surface shape of object to be detected. Moreover, it also depends on micro-shape of those defects. But, defects are random with regard to shape, major axis direction, size, and location. So, based on discussion above illumination configures is determined by combining simulation and experiment.

9.6.2.1 The Area Ratio of Saturation Pixel

First, the specular region is comprised of those pixels whose intensity is close to saturation value. It is because that the object such as chrome-plated surface is very strong reflective surface. That is,

$$p = \begin{cases} \text{saturation pixal,} & \text{if } I_p > T_s \\ \text{other,} \end{cases} \tag{9.44}$$

where I_p is the intensity of pixel p, and T_s is a threshold. For an 8-bit gray image, by some experiments it is reasonable that the threshold T_s is defined as 0.95×255 in our application. The area's ratio is defined as ration of the total area of specular regions to the area of the whole inspection surface. So, the area ratio σ_s can be described as $\sigma_s = N_s/N$.

9.6.2.2 The Defect Expressivity

The defect expressivity σ_d is defined by the follow formulations. The mean intensity of the region of interest:

$$M = \sum_{i=0}^{255} \left(i \times \frac{n_i}{N} \right), \tag{9.45}$$

the mean intensity of the region of defect:

$$M' = \sum_{i=0}^{255} \left(i \times \frac{n_i}{N'} \right), \tag{9.46}$$

where n_i is the number of pixels whose intensity is i, N is the total number of the region of interest, N' is the defect region. So, the defect expressivity σ_d is defined in Eq. (9.47).

$$\sigma_d = \frac{M' - M}{M}, \tag{9.47}$$

The bigger the defect expressivity σ_d, the more easily defects can be detected. The defect expressivity depends on the reflection properties of the surface under investigation, the illumination parameters, and the defect shape and direction.

9.6.3 Defect Inspection on Cover-Knobs

A complete inspection is defined as a series of consecutive inspection processes starting from image sampling, image analysis to finally result in the display and output. So, a complete inspection procedure includes:

1. Sample analysis. Analysis reflection properties and geometry data of sample to design inspection method;
2. Set system parameters. (system calibration, the inspection guidelines);
3. Inspect defects;
4. Display or output the inspection result.

9.6.3.1 Setting the Inspection System Parameters

The cover-knob, as shown in Fig. 9.1, is covered with chrome-plated ABS resin [20], which is one of the shiniest specular reflection surfaces and is broadly found in the automotive industry because of their bright appearance. Figure 9.44, respectively, shows four images of four surfaces of the object. Some defects called scratch have been detected by the human operator. As shown in Fig. 9.44b, these defects appear on one side surface of the surface. On other surfaces, there are not defects. It is worth described that we need to inspect whole object by means of inspecting its four surfaces, respectively, due to the object is complicate. The next inspection only is performed by the surface with defects. Others surface inspections are finished by the same way.

Refer to paper [57, 58] and documents from the inspection guideline, its reflection properties are as follows: specular reflectance $\rho_s = 0.9$, diffuse reflectance $\rho_d = 0.1$, surface roughness $\sigma = 0.07$. So, the camera parameters are determined according to the following constraints. (1) Defect size. The part with defect which size is more than 1 mm will be rejected. This is come from the guideline of the object. (2) Object real size. Concerning real size of the object, one only take into account its maxis dimensions length and width, which, respectively, are $l_{max} = 70$ mm, $w_{max} = 35$ mm. So, according to our defect inspection algorithm and Eq. (9.43), the resolution of the system is 0.33 mm/pixel. To satisfy the defect dimensions requirements, we need to perform the camera calibration. The camera and the illumination system parameters are shown in Table 9.3.

1. Set the object pose.
   Figure 9.41 shows 3D plot of the cover-knob. One face is cut from the object. The object is almost symmetrical around X–Z plane, so it is unnecessary to investigate the difference between various R_x. In addition, the change around z-axis can be described by using of changing illumination azimuth. So, one further and only discuss the image how to change when the object is rotated

Table 9.3 System parameters for cover-knob inspection

Parameter	Value	Parameter	Value
Image size	640 × 480	Light number	4
Special resolution	0.33 mm/pixel	Illumination intensity	15 W
Mini defect	1 mm	Illumination evaluation angle	30°
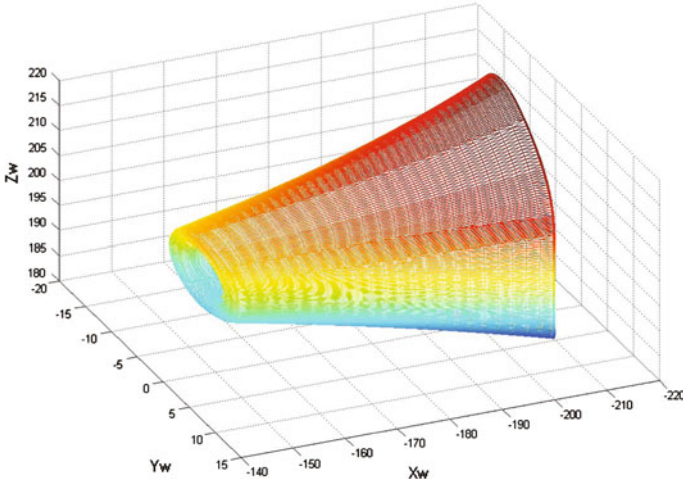
Fig. 9.41 3D plot of the cover-knob

around *y*-axis. A set of images are generated by the numerical platform proposed in Sect. 9.3. The object pose is represented by (R_x, R_y, R_z) and illumination direction (θ, φ). Based on simulation by using the numerical platform shown in Sect. 9.3, object is fixed at $(0°, -8°, 0°)$, and illumination directions (θ, φ) are set as $(30°, 45°)$, $(30°, 135°)$, $(30°, 225°)$, and $(30°, 315°)$.

2. Determine illumination direction and intensity.

   The illumination elevation angle and illumination intensity is discussed by some real object with defects after discussion about the object pose and the illumination azimuth. Under the same position and same illumination intensity, the defect expressivity [59] when the elevation angle is 30° is bigger than the one when the elevation angle is 60°. On the other hand, when the elevation angle is fixed at 30°, the defect expressivity when illumination intensity is 15 is bigger than the one when illumination intensity is 5. So, for the object, the illumination intensity is 15 and the elevation angle is determined as 30°.

9.6.3.2 Inspection Results on a Cover-Knob

Figure 9.42 shows four original gray images captured from the work piece illuminated by four different directions in turn. In each original image shown in Fig. 9.42, most of pixels are very dark, whose intensities are close to 0. It confirmed that the illumination system avoids specular reflection from the flawless surface and the edges also appear as high contrast. Some bright spots which are close to saturation appear in these original images, as shown in Fig. 9.42. So, we can indicate that the light system with side illumination success in reducing the reflection from smooth surface, although some bright appears on the edge.

(a) $(30^0, 45^0)$

(b) $(30^0, 135^0)$

(c) $(30^0, 225^0)$

(d) $(30^0, 315^0)$
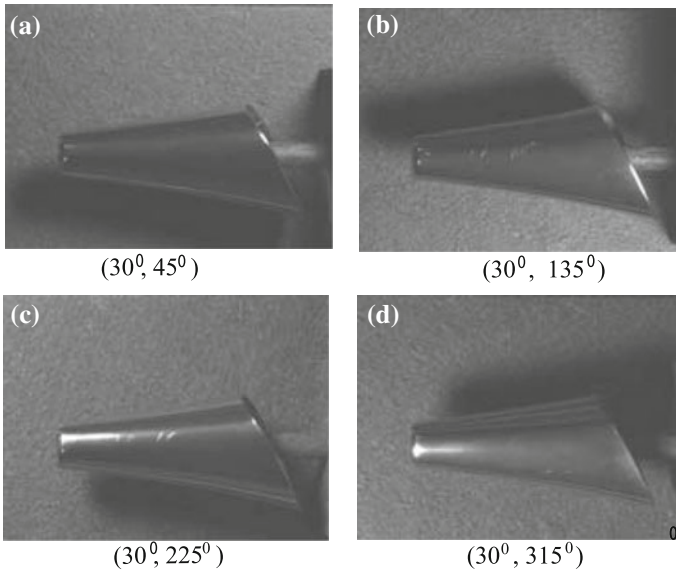
Fig. 9.42 Four original images of cover-knob inspection

On the other hand, the shape and location of defects are random and complicated, so some defects may not be illuminated under some illumination conditions. As shown in the image shown in Fig. 9.42a, d, the defects are not clear. This case results from the relation between the defects direction and the illumination direction. In contrast, in the image shown in Fig. 9.42b, c, the defects appear bright reflection. In particular in Fig. 9.42c, the defects appear very high contrast with their around region.

Fig. 9.43 The synthetic image of the car cover-knob

Fig. 9.44 All defects of the car cover-knob

From the four images, a synthetic image is reconstructed by the multi-image reconstruction method shown in Sect. 9.4, as shown in Fig. 9.43. From the synthetic image, we use the defect inspection method shown in Sect. 9.5 to find all defects shown in Fig. 9.44. Through this experiment, we can conclude the following conclusion. First, proper system parameters can significantly improve the image quality. In particular, the illumination direction plays a very important role to increase the defect expressibility. Second, a single image reconstructed from a set of images is used to integrate information and eliminate specular reflection effect. Because we cannot know the defect direction or shape in advance, only various images illuminated from different direction can ensure defect information appear in an image among the set of image. Third, generally speaking, for a reasonable complexity surface, four images captured from different direction are enough. For example, the defects in the cover-knob are scratches which master direction is along the 45° from the x-axis, so the defects are not clear in Fig. 9.42a, d. In contrast, the defects are very clear (b) and (d).

9.6.4 Defect Inspection on Car Doorknobs

9.6.4.1 Defects Analysis in Image from a Car Doorknob

The car doorknob, whose size is of 180×40 mm, has been inspected using the proposed inspect system. On the sample, some defects have been found by manually. According their shape, these defects generally can be classified into two classes: dot and scratch. The appearance of scratch often is thin and long. In particular, these scratch defects with various orientations exist on different locations on the surface. According to the method proposed in this dissertation, the system parameters are first set. Under these conditions, a set of images are captured from
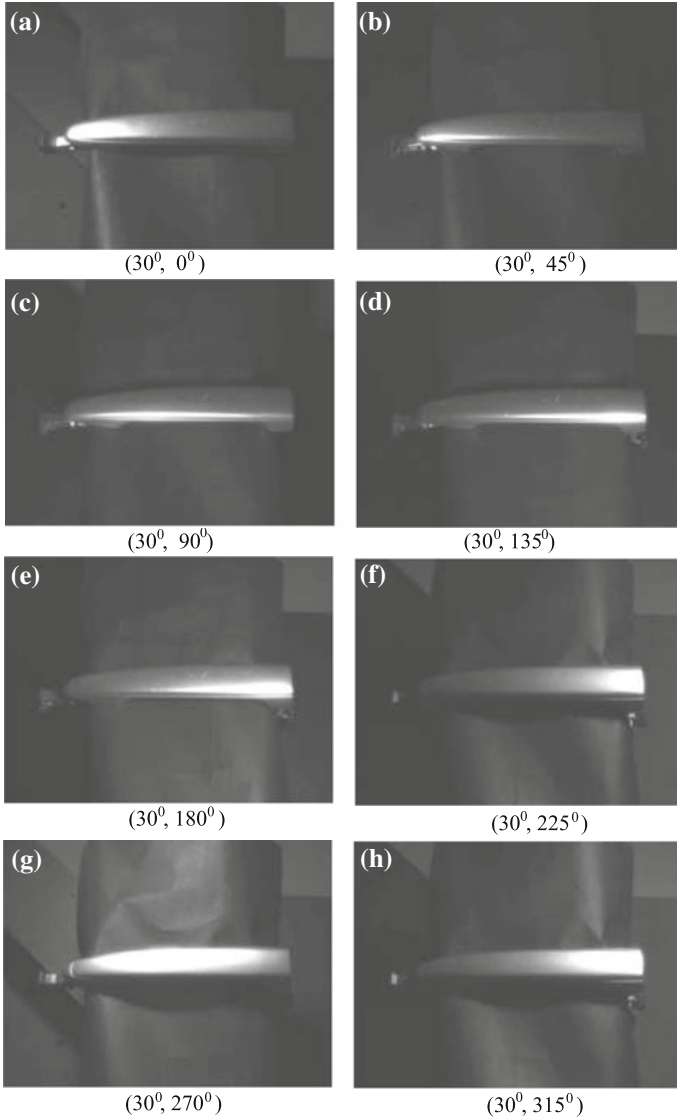
Fig. 9.45 Eight original gray images of the car doorknob

the car doorknob. Figure 9.45 shows eight original images of the car doorknob and their corresponding illumination directions. From these original images captured from different illumination directions, the following case can be found:

1. Some defects are covered by bright areas.
   As shown in Fig. 9.45a, a bright area appears in the left part of the object surface image. It is obvious that defects in the area cannot be found; however, method is

performed. It is because that intensity of all pixels of this area is saturated; that is, defects are covered by this bright area. Likewise, right defects on the surface also are covered by the right bright region, as shown in Fig. 9.45f. So, it is impossible to find all defects in the image although defects on other region of the image can be detected by traditional method. Collection or integrating all defects from various images is the only approach. The method proposed in this dissertation is to integrate all defects, which are visible in different image, in a single image.

2. Defect expressivity is a relation to the defect direction (especially for scratch) and illumination direction.

According to our definition in Sect. 9.5, the defects with higher defect expressivity would be detected more easily. As shown in Fig. 9.45e, a scratch is horizontal and very clear. However, the defect almost cannot be found by the

Fig. 9.46 The synthetic image from images shown in Fig. 9.45a, c, e, g

Fig. 9.47 The synthetic image from images shown in Fig. 9.45b, d, f, h

naked eye from the Fig. 9.45a. Likewise, the case also occurs in other defects and other images.

Figure 9.46 shows the synthetic image reconstructed from four images, which are shown in Fig. 9.45a, c, e, g. Comparing the synthetic image and an original image, for example Fig. 9.45g, defects located in the right region of the object surface appear in reasonable high contrast. On the other hand, defects located in the left are not very clear. In contrast, in another image shown in Fig. 9.48, some defects also are not clear. However, Fig. 9.47 is reconstructed from four images, which are shown in Fig. 9.45b, d, f, h.

9.6.4.2   An Improvement Approach for Defect Inspection

So, an improvement method is proposed to ensure completely and robustly detect all defects. It aims to improve the quality of the synthetic image from which all defects can be robustly detected. Figure 6.15 shows the improved approach for defect inspection.

First, a synthetic image is reconstructed from four images captured from the object surface illuminated by four different directions. The synthetic image is called as I_{syn}^1. As shown in Fig. 9.48, four images are I_1, I_3, I_5, I_7 captured under four illuminations L_1, L_3, L_5, L_7. In generally, four illumination directions are same elevation angle with 90° interval azimuth. The reconstruction method has been described in Sect. 9.4. Likewise, another synthetic image I_{syn}^2 is obtained from other four images, which also are captured from four various illumination directions. As shown in Fig. 9.48, four images are I_2, I_4, I_6, I_8 captured under four illuminations L_2, L_4, L_6, L_8. In general, four illumination directions are of same elevation angle with 90° interval azimuth. In addition, I_2, I_4, I_6, I_8 are, respectively, 45° more than I_1, I_3, I_5, I_7. To the end, $I_{[syn]}^2$ and I_{syn}^2 are fused to a single image I_{syn} by the Eq. (6.8).
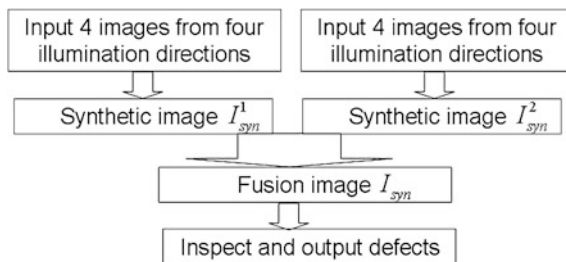
Fig. 9.48 Improved approach for defect inspection on curved surface

Fig. 9.49 The synthetic image of the car doorknob by the improved method
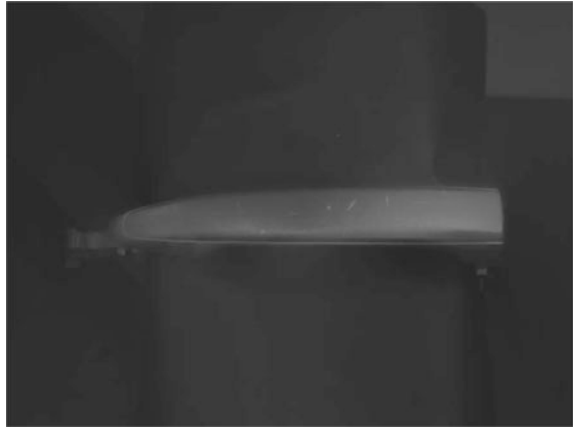
Fig. 9.50 All inspected defects of the car doorknob

Table 9.4 Features of all defects on a car doorknob

Defect	Center	Area (pixel)	Perimeter (pixel)	Compactness
1	151, 248	58	28	0.936
2	191, 234	50	24	1.009
3	212, 238	61	29	0.933
4	264, 232	111	43	0.852
5	151, 248	96	42	0.822
6	191, 234	63	27	1.026
7	212, 238	167	53	0.853
8	264, 232	102	40	0.884
9	151, 248	125	52	0.758
10	191, 234	49	24	0.999

$$I_{\text{syn}} = c_1 I_{\text{syn}}^1 + c_2 I_{\text{syn}}^2$$
$$c_1 + c_2 = 1, \tag{9.48}$$

where c_1 and c_2 are weight coefficients. In practical application, the simplest algorithm is adopted, that is I_{syn} is the average of I_{syn}^1 and I_{syn}^2.

The final synthetic image I_{syn} of the car doorknob is shown in Fig. 9.49. By comparing image I_{syn} with image I_{syn}^1 and I_{syn}^2, it is obvious that the improved approach is more robust. It can completely ensure all defects with various shape, direction, and location can appear in a single image. Moreover, all defects appear even contrast. This case is favor of the next defect extraction. Figure 9.50 shows defect inspection results obtained from Fig. 9.49 by using the improved method. And the Table 9.4 shows various features of all defects. The inspection result is consistent with the manual inspection result. Furthermore, the other experiments also obtained nice results. Therefore, one can conclude that this inspection system can reliably detect defects on HSR curved surfaces and it is robust to the shape and location of the defects. Table 9.4 shows some features of all defects on the car doorknob.

9.6.5 Summary

In this chapter, a new vision defect inspection system is developed, and its application on HSR curved surface also is presented. In addition, two typical examples of kinds of experiments are presented and discussed.

In the first experiment, the object called cover-knob is a chrome-plated surface [20], which acts like a mirror. In particular, in the second experiment on the car doorknob, two group images are captured to reconstruct an image in which all defects are successfully integrated. Some conclusions can be summarized as follows:

1. The inspection system is comprised of an acquisition system, an inspection platform, and a host PC. A set of images are captured from the image acquisition system when the object is illuminated by a group LED light source in turn. The LED light source is adapted to greatly reduce the effect of the specular reflections. In addition, the light source that illuminated the HSR curved surface from a concern direction ensures defects are efficiently illuminated.
2. The object pose and illumination direction and intensity play very important role in inspection system. The object pose and illumination direction and intensity are determined by analyzing numerical simulation and real image. In particular, the area ratio of saturation pixel is adapted to determine the object pose in the CAD-guided simulation platform, and the illumination parameters are determined by computing the defect expressivity from real object samples. Real

applications confirm that proper system parameters are very important to achieve defect inspection.

3. The vision inspection system is feasible and robust to achieve defect inspection on curved surface with highly specular reflection. A synthetic image reconstructed from a set of images appears more uniform gray change. In particular, all defects successfully integrate in the synthetic image. So, all defects have been successfully detected from the single image. For a reasonable complexity surface, defect inspection can be robustly performed from four images. For more complicated curved surface, bright area cannot be completely avoided. More bright areas cover defects or defects are not be illuminated because that defects are of various shapes, locations, and directions. An improved approach was proposed to achieve defect inspection on more complicated curved surface. In the improved approach, two synthetic images is, respectively, reconstructed from four images. And then a synthetic image is fused from the two synthetic images. The vision inspection system is robust although the defects are of various shapes, directions, and locations are random. Lots of experiments show that defect inspection is feasible and robust to achieve defect inspection on curved surface with highly specular reflection.

9.7 Concluding and Remarks

In this chapter, a new vision defect inspection system has been shown. In particular, a synthetic is reconstructed from a set of images captured under various illumination conditions (e.g., various illumination directions). This vision defect inspection is very robust for HSR curved surfaces, especially for chrome-plated surfaces. The reconstruction algorithm efficiently eliminates the trouble from specular reflection, which always makes traditional defect inspection method error. Meanwhile, an efficient and robust method is proposed to extract all defects from a single synthetic image, in which all defects are integrated by the proposed reconstruction algorithm.

In Sect. 9.2, some basis theory background of vision image-based defect inspection approach on highly specular surface was presented. By analyzing the reflection from HSR curved surface, it can be concluded that the reflection from defect region will distinctly different from flawless surface when the image is captured under some conditions involving view sensor, illumination, and so on. This difference is very enormous for highly specular reflection surface while very weak diffuse reflection.

In Sect. 9.3, a CAD-guided numerical simulation platform has been created. Simulated images can significantly reduce the engineering time to design and successfully implement a vision system. It confirms that proper illumination and view direction can help to realize defect inspection on flat specular surface. However, due to the complex reflection from perfect curve surface or random

defects, from a single image it is almost impossible to extract all defects. Another important reason is that the strong specular reflection maybe lead defects covered in highlight regions and the weak diffuse must lead defect appear very low contrast with perfect region. By using the platform, system parameters (light and view parameters) and object pose can be analyzed to ensure least specular reflection. Moreover, defect inspection and machine vision have been designed in the CAD-guided numerical simulation platform. And the result can be applied to design special inspection approach according to the special task for HSR surface. So, in the next section, a new novel method is developed to integrate all defects in a single image and eliminate trouble from specular reflection.

Section 9.4 describes our basic algorithm to integrate all defects in a single image from a set of image and process specular reflection resulted from highly specular surface. A synthetic image is reconstructed from the gradient field, which is obtained from a set of images captured under different conditions. After this processing, better image was produced to realize defect inspection on curve surface with highly specular reflection surface.

In Sect. 9.5, a new defect segmentation technique, which combines the morphology technique and template matching, is proposed to achieve defects inspection from the synthetic image reconstructed by the method shown in Sect. 9.4.

In Sect. 9.6, a vision defect inspection system is developed. Image acquisition method is especially determined to obtain less specular reflection and high defect expressivity. In addition, some defect inspection experiments on HSR curved surface are successfully performed by the proposed system. A synthetic image reconstructed from a set of images appears more uniform gray change. In particular, all defects successfully integrate in the synthetic image. So, all defects have been successfully detected from the single image. For a reasonable complexity surface, defect inspection can be robustly performed from four images. For more complicated curved surface, bright area cannot be completely avoided. More bright areas cover defects or defects are not be illuminated because that defects are various shapes, locations, and directions. An improved approach was shown to achieve defect inspection on more complicated curved surface. In the improved approach, two synthetic images are, respectively, reconstructed from four images. And then a synthetic image is fused from the two synthetic images. The vision inspection system is robust although the defects are various shapes, directions, and locations are random. Lots of experiments show that defect inspection is feasible and robust to achieve defect inspection on curved surface with highly specular reflection.

In the future, one would like to focus on increasing the performance of this inspection system and implement the inspection system on the production line. However, deriving the optimal illumination system from the surface topography and reflection properties of the surface under investigation is the next work to be done. In particular, for more complicate surface, segmenting the 3D surface [60–62] into several different regions and setting the corresponding inspection system parameters would make this system high performance.

References

1. Newman T, Jain A (1995) A survey of automated visual inspection. Comput Vis Image Underst 61:231–262
2. Sanderson C, Weiss LE, Nayar SK (1988) Structured highlight inspection of specular surfaces. IEEE Trans Pattern Anal Mach Intell 10(1):44–55
3. Bakolias C, Forrest AK (1997) Dark field scheimflug imaging for surface inspection. SPIE Proc Mach Vis Appl Ind Inspection V 3029:57–68
4. Malamas E, Petrakis E, Zervakis M, Petit L, Legat JD (2003) A survey on industrial vision systems, applications and tools. Image Vis Comput 21:171–188
5. Kim CW, Koivo AJ (1994) Hierarchical classification of surface defects on dusty wood boards. Pattern Recog Lett 15(7):713–721
6. Piironen T (1990) Automated visual inspection of rolled metal surfaces. Mach Vision Appl 3 (4):247–254
7. Fernandez C, Platero C, Campoy P, Aracil R (1993) Vision system for on-line surface inspection in aluminum casting process. In: Proceedings of the IEEE International Conference on Industrial Electronics, Control, Instrumentation and Automation, Maui, HA, pp 1854–1859
8. Rohrmus D (1999) Invariant texture features for web defect detection and classification. In: SPIE Proceeding of machine vision systems for inspection and metrology VIII, vol. 3836, Boston, MA, 144–155
9. Boukouvalas C, Kittler J, Marik R, Petrou M (1994) Automatic grading of ceramic tiles using machine vision, In: Proceedings of the IEEE international symposium on industrial electronics, pp 13–18
10. Nayar SK, Sanderson AC, Weiss LE, Simon DA (1990) Specular surface inspection using structured highlight and Gaussian images. IEEE Trans Robot Autom 6(2):208–218
11. Caulier Y, Bourennane S (2008) An image content description technique for the inspection of specular objects. EURASIP J Adv Signal Process vol. 2008, Article ID 195263
12. Caulier Y, Spinnler K Arnold M (2009) Automatic detection of surface and structural defects on reflection workpieces. Photonink Int online
13. Caulier Y, Spinnler K, Bourennane S, Wittenberg T (2008) New structured illumination technique for the inspection of high-reflective surfaces: application for the detection of structural defects without any calibration procedures, EURASIP J Image Video Process vol. 2008, Article ID 237459
14. Petz M, Tutsch R (2002) Optical 3D measurement of reflecting free formed surfaces. Technical report, Institut fur Produktions-messtechnik-iprom, Technische Universit at Braunschweig, Braunschweig, Germany
15. Jordan A, Kahmann F, Luecking F (2003) Dual solution for high-end stripe inspection. Mater Test 26(4):66–71
16. Leon FP, Beyerer J (1997) Active vision and sensor fusion for inspection of metallic surfaces. In: Proceedings of the SPIE intelligent robots and computer vision XVI: Algorithms, Techniques, Active Vision, and Materials Handling, vol. 3208, Pittsburg, Pa, USA, 394–405
17. Aluze D, Merienne F, Dumont C, Gorria P (2002) Vision system for defect imaging, detection, and characterization on a specular surface of a 3D object. Image Vis Comput 20:569–580
18. Qu XH, Hu Y, Han F (2003) An automatic inspection system for random defects on strongly reflective and complex surface. Opto-Electron Eng 30(2):32–35
19. Badger JC, Enright ST (1996) Auomatic surface defect inspection system. Iron Steel Eng 3:48–51
20. Zhang Z, Ren SQ, Miyake T, Fujiwara H, Imamura T (2008) Processing reflections on metallic surfaces using a special random sampling method. Int J Innovative Comput Inf Control 4(7):1595–1606

21. Ren SQ, Zhang Z, Miyake T, Fujiwara H, Imamura T (2008) Edge detection of highly specular surface using special random sampling method. Proceedings of international conference on innovative computing, information and control (ICICIC2009), pp 1–4
22. Torrance K, Sparrow E (1967) Theory for off-specular reflection from roughened surfaces. J Opt Soc Am 57:1105–1114
23. Ward GJ (1992) Measuring and modeling anisotropic reflection. Proceedings of the 19th annual conference on computer graphics and interactive techniques, SIGGRAPH'92, pp 265–272
24. Horn BKP (1970) Shape from shading: a method for obtaining the shape of a smooth opaque object from one view. MIT Project MAC Int. Rep. TR-79 and MIT AI Lab. Technical Report p 232
25. Horn BKP (1981) Hill shading and the reflectance map. Proc IEEE 69(11):14–47
26. Woodham RJ (1978) Photometric stereo: a reflectance map technique for determining surface orientation from image intensity. Proc SPIE 155:136–143
27. Lambert JH (1960) Photometria sive de mensura de gratibus luminis, colorum et umbrae, Augsberg. Eberhard Klett, Germany
28. Phong B (1975) Illumination for computer generated pictures. Communication ACM 18:311–317
29. Ikeuchi K (1981) Determining surface orientations of specular surfaces by using the photometric stereo method. IEEE Trans Pattern Anal Mach Intell 3(6):661–669
30. Sanderson AC, Weiss LE, Nayar SK (1988) Structured highlight inspection of specular surfaces. IEEE Trans Pattern Anal Mach Intell 10(1):44–55
31. Healey G, Binford T (1988) Local shape from specularity. Comput Vis Graph Image Process pp 62–86
32. Nayar SK, Ikeuchi K, Kanade T (1991) Surface reflection: physics and geometrical perspectives. IEEE Trans Pattern Anal Mach Intell 13(7):611–634
33. Pentland A (1984) Local shading analysis. PAMI 6:170–187
34. Tsai Y, Roger A (1987) Versatile camera calibration technique for high-accuracy 3-D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE J Robot Auto, RA-3(4)
35. Barsky S, Petrou M (2003) The 4-source photometric stereo technique for three dimensional surfaces in the presence of highlights and shadows. IEEE Trans Pattern Anal Mach Intell 25(10):1239–1252
36. Zang Z, Ren SQ, Miyake T, Fujiwara H, Imamura T, Hara K, Nishino K, Ikeuchi K (2003) Determining reflectance and light position from a single image without distant illumination assumption. Process Ninth IEEE Int Conf Comput Vis 2003:100–107
37. Woodham RJ (1994) Gradient and curvature from the photometric-stereo method, including local confidence estimation. J Opt Soc Am 11(11):3050–3068
38. Wolff LB, Boult T (1991) Constraining object features using polarization reflectance model. IEEE Trans Pattern Anal Mach Intell 13(7):635–657
39. Nayar SK, Fang XS, Boult T (1996) Separation of reflection components using color and polarization. Int J Comput Vis 21(3):163–186
40. Sato Y, Ikeuchi K (1994) Temporal-color space analysis of reflection. J Opt Soc Am A 11(11):2990–3002
41. Robby TT, Ikeuchi K (2005) Separating reflection components of textured surfaces using a single image. IEEE Trans Pattern Anal Mach Intell 27(2):178–193
42. Mukaigawa Y, Miyaki H (2001) Photometric image-based rendering for image generation in arbitrary illumination. Proc Int Conf Comput Vis pp 643–649 (2001)
43. Lischinski FD (2002) Gradient domain high dynamic range compression. Proc SIGGRAPH 2002:249–256
44. Daubechies I (1988) Orthonormal bases of compactly supported wavelets. Commun Pure Appl Math 41(7):909–996
45. Coe S (2000) Automatic tile inspection, Surface Inspection Limited, International Ceramics, Bristol, U.K., Issue 1

46. Glazer A, Sipper M (2008) Evolving an automatic defect classification tool. Lecture notes in computer science, in Proc. Applications. Evolutionary Computing- EvoWorkshops 2008, 194–203
47. Wang CH, Kuo W, Bensmail H (2006) Detection and classification of defect patterns on semiconductor wafers. IEEE Trans 38:1059–1068
48. Lee SH et al (2005) Automatic defect classification using boosting. Proc IEEE Comput Soc Mach Learn Appl pp 15–17
49. Kameyama K, Kosugi Y (1999) Semiconductor defect classification using hyperellipsoid clustering neural networks and model switching. Proceedings of IEEE international joint conference neural networks, pp 3505–3510
50. Jeong YS, Kim SJ, Jeong MK (2008) Automatic identification of defect patterns in semiconductor wafer maps using spatial correlogram and dynamic time warping. IEEE Trans Semicond Manuf 21(4):625–637
51. Liu HX, Zhou W, Kuang QW (2010) Defect detection of ic wafer based on spectral subtraction. IEEE Trans Semicond Manuf 23(1)
52. Alin H et al (2008) Study on defect detection of IC wafer based on morphology. Proc SPIE-Int Soc Opt Eng, Electron Imaging Mult Technol V:334–337
53. Tsai DM, Yang CH (2005) A quantile-quantile plot based pattern matching for defect detection. Pattern Recogn Lett 26:1948–1962
54. http://www.aegis-elec.com/products/Teli-CSFV90BC3.html
55. Telljohann A (2008) Introduction to building a machine vision inspection, handbook of machine vision. Wiley-VCH Verlag GmbH, Hornberg
56. Jahr I (2008) Lighting in machine vision handbook of machine vision. Wiley-VCH Verlag GmbH, Hornberg
57. Rump M, Muller G, Sarlette R, Koch D, Klein R (2008) Photo-realistic rendering of metallic car paint from image-based measurements. Eurographics 27(2):2008
58. Gunther J, Chen TB, Goesele M, Wald I, Seidel HP (2005) Efficient acquisition and realistic rendering of car paint. Proceedings of VMV 2005, Erlangen, Germany
59. Li CJ, Zhang Z (2001) A new illumination system for defect inspection on highly specular surface Proceedings of the 2011 international conference on signal and information processing (ICSIP2011), pp 569–576
60. Besl PJ, Jain RC (1988) Segmentation through variable-order surface fitting. IEEE Trans Pattern Anal Mach Intell 10(2):167–192
61. Cohen EH, Singh M (2007) Geometric determinants of shape segmentation: tests using segment identification. Vision Res 47:2825–2840
62. Yamazaki I, Natarajan V, Bai ZJ, Hamann B (2010) Segmenting point-sampled surfaces. Vis Comput 2010(26):1421–1433
63. Pernkopf F, Leary Paul O (2002) Visual inspection of machined metallic high-pecision surfaces. EURASIP J Appl Signal Process 2002(7):667–678

Part II
Applications and System Integration for Vision-Based Inspection

Chapter 10
Robotic Inspection Systems

Christian Eitzinger, Sebastian Zambal and Petra Thanner

Abstract Industrial quality control often includes the inspection of parts of complex geometry. While such an inspection can be quite easily done by humans, it poses certain challenges if the task is to be automated. Quite often, robots are used for handling the part to acquire a large number of images, each showing a certain area of the surface. The process of acquiring sequences of multiple images also has implications for the machine vision and analysis methods used in such tasks. This chapter covers all topics that relate to the implementation of robotic inspection systems for industrial quality control. The focus is on machine vision, while aspects that deal with robotics will only be addressed at a conceptual level.

Contents

C. Eitzinger (✉) · S. Zambal · P. Thanner
Profactor GmbH, Im Stadtgut A2, 4407 Steyr-Gleink, Austria
e-mail: christian.eitzinger@profactor.at

10.1 Introduction

Industrial quality control is often realized as an automated end-of-line inspection system that checks certain features on the products. Inspection covers functional aspects of the product, such as whether all components are mounted on an assembly, as well as aesthetic properties, such as checking for scratches on a polished surface. We would like to make a clear distinction between *inspection* and *measurement*. Measurement systems provide quantitative information about a product in the form of certain physical units that can be traced back to internationally accepted standards. Measurement systems usually aim at achieving high precision. Inspection systems generate qualitative information about a product, usually by assigning the product to a certain class, e.g. "accept", "reject" or "rework". The main property of an inspection system is its classification accuracy, often represented in the form of a confusion matrix.

Robotic inspection systems use robots to perform the inspection task. The term "inspection robot" is often used in conjunction with mobile robots used for the inspection of tunnels, channels or pipes [1], sometimes with highly specific kinematics [2]. In [3], a standard industrial robot is used for quality control of printed circuit boards. However, in this application, the robot is only used to handle the part and does not directly support the inspection process. The kind of inspection robots that we will be investigating in this chapter is based on standard industrial robots that guide a vision-based sensor system over a part of complex geometry, such as those described in Woern et al. [4]. In [5], the idea is extended with a focus on selectively reworking of defective areas on the part. The common aspect of visual inspection with robots is that multiple images of a single part are required, where each image provides an optimal view of an area to be inspected on the part. The key elements of optimizing the viewpoints of a vision sensor are covered in Yang and Ciarallo [6] and in the references mentioned there. The reason for choosing a robot over using multiple cameras—which is often the cheaper solution—is the higher flexibility of the robot and the possibility of adapting the inspection process to different product variants. Depending on the application, the robot may handle either the image acquisition system, including the camera and the illumination [7], or the product itself. Within the class of robotic inspection systems, we also want to distinguish between two different types of inspection tasks: in the first, simpler case, the robot is positioning the camera at a sequence of pre-defined positions and an image is acquired from that position. Typically, the number of positions is low, e.g. up to 15 different locations. Each image is processed separately and the processing steps and parameters are set up manually. A typical application is completeness inspection of a complex assembly. The inspection system has to check whether the assembly contains all components, whether the correct type of component is mounted on the assembly and whether the components are in their correct positions. An example is shown in Fig. 10.1.

A typical scenario is the inspection of a fully assembled car engine, where the correct mounting of several plugs, pumps, hoses and cables needs to be verified.
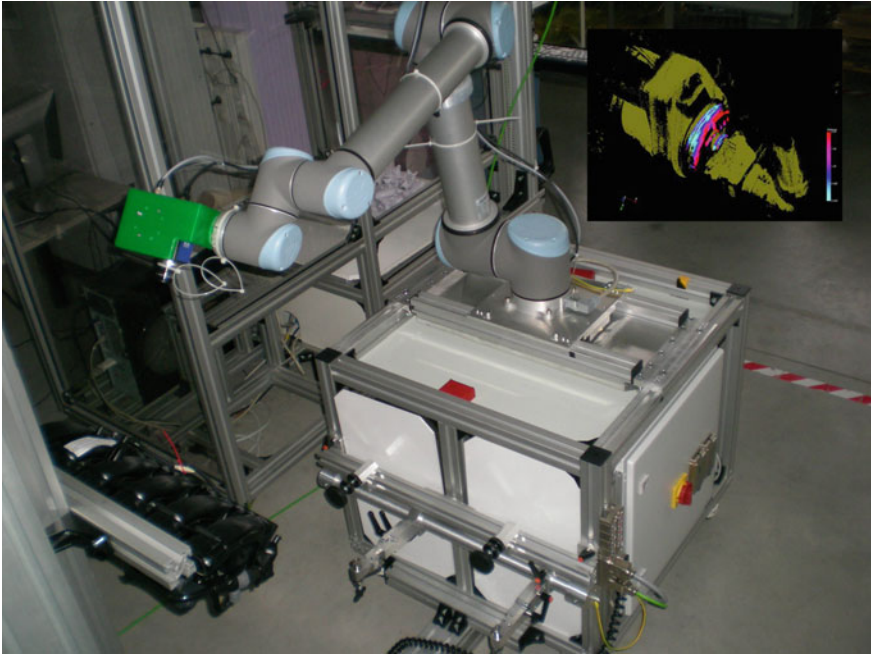
Fig. 10.1 An inspection robot for completeness inspection in a laboratory setting. The insert shows a point cloud of the part with the false component *highlighted in red*

Obviously, there is no single viewpoint from which all these components can be seen, and therefore, the camera needs to be positioned to obtain an optimal view of each component. In this case, the inspection task is—at a conceptual level— identical to standard machine vision systems. Single images are acquired, they are analysed independently, and a decision is made based on the results of each single image.

In this chapter, however, we want to deal with the second, more complex case, where the robot needs to inspect a larger area by moving a camera over the whole surface. In this case, a very large number of overlapping images are acquired, typically several hundreds of images. These images cannot be processed independently, and manually setting up the inspection process is impossible, simply due to the large number of images. The typical application is surface inspection, where defects have to be found on the surface of a part. One such scenario is the inspection of chrome-plated fittings for small scratches and dents. The shiny surface allows an inspection only for a specific geometrical arrangement of illumination, camera and surface, so that a 3D free-form surface can only be fully covered by splitting it into a large number of regions and finding an optimal viewpoint for each of these regions. Due to the large number of regions, the inspection process cannot be done in a stop-and-go motion, but needs to be done by continuously moving the camera across the surface and adjusting its position in all six degrees of freedom.

This kind of inspection process poses several challenges that will be addressed in the remaining parts of this chapter:

- The large number of images does not allow a manual set-up of the whole inspection process. Many of the parameterization steps need to be done automatically, such as identifying the regions in the image that are of sufficient quality to be used for processing.
- The robot's path cannot be generated by a teaching process, because it is impossible for a human teacher to maintain an accurate distance and angle to the local surface normal. Instead, a process model of the image acquisition process is required that provides the basis for automatic planning of the robot's path.
- In many advanced inspection processes, several images of the same location are required, for example with different shutter times to extend the dynamic range of the images. To merge such images, accurate registration is needed, which in many cases requires fairly accurate synchronization between the robot's motion and the image acquisition process.

In the following sections, we will first describe how these challenges are addressed at a conceptual level and then proceed by presenting the details of two different, typical realizations of this concept.

10.2 General Concept and System Overview

A robotic inspection system is a complex technical system that requires a close interaction between its single components. The initial set-up of an inspection task for a 3D free-form surface cannot be done manually, because a human operator cannot define with sufficient accuracy the path that the robot has to follow. In many cases, the set-up of the image acquisition system relies on a specific geometrical configuration between camera, illumination and surface. This configuration only allows minimal tolerances, and consequently, manual positioning will not be an option. Therefore, robotic inspection systems need a (semi-)automatic offline configuration process that is used to define the whole inspection task so that it can afterwards be executed fully automatically. This includes elements such as 3D CAD models of the part, a process model of the image acquisition process, algorithms for path planning and calibration methods. All of these components will be further described in Sect. 10.2.1.

Once the offline process is finished and the inspection task has been defined, the robot should be able to automatically execute the inspection process, which may also include compensation for small deviations that may occur, e.g. in the positioning of the product to be inspected. This online inspection process includes the standard machine vision elements, such as low-level image processing, image segmentation, feature extraction, classification and decision-making. For robotic inspection systems, additional elements are needed, such as synchronization

between image acquisition and the robot's position, projection of defects from the 2D images to the 3D surface and identifying areas in the image that are suitable for inspection. These topics will be addressed in Sect. 10.2.2.

10.2.1 Offline Components

Before the automatic inspection can be started, the inspection process has to be set up for the particular product variant. While in many machine vision tasks this set-up can be done manually, the robotic inspection tasks that we will discuss in the following sections require a semi-automatic set-up process because of the complexity of the task. This process is shown in Fig. 10.2.

The usual starting point is a 3D CAD model of the object to be inspected. This model needs to be stripped of all elements that do not include shape information such as measurements and other metadata. In many cases, the CAD model is converted into a simple standard format to remove the complexity of 3D free-form shapes that may be present. This can be achieved either by point clouds or preferably by a triangular mesh. However, shape information alone is not sufficient, and it needs to be augmented with additional information that is not typically included in CAD formats [8].

Such information may include, e.g., the inspection parameters, such as allowable defect sizes per defect type or the acceptable maximum number of defects. These parameters may be different for different areas on the product. Therefore, the CAD model sometimes needs to be split into areas to which different quality criteria apply. Occasionally, additional information will also include surface properties, such as whether the surface is milled, sand-blasted or polished. This can be used to switch image acquisition parameters during the acquisition process.

Once these augmented CAD data are available, we can use them to generate a path for the robot. This path has to fulfil several criteria:
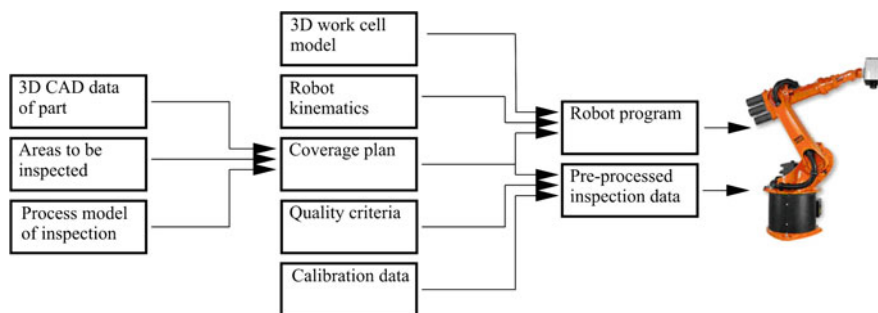
Fig. 10.2 Offline processing: the components and their interaction

- The path has to cover all surface elements that need to be inspected. For many inspection tasks, several images of a single surface element need to be acquired, e.g. under different illumination or with different shutter times. Therefore, each surface element has to be acquired multiple times, while the robot is moving.
- The path has to comply with certain time constraints, e.g. the speed may not be so fast that the maximum frame rate of the camera is exceeded, or if the frame rate cannot be adapted, the robot's speed has to be kept within quite tight tolerances.
- All the positions along the path that the robot is expected to follow must be reachable for the robot, i.e. they must be within the workspace of the robot.
- The whole path of the robot must be free of collisions, i.e. the robot shall not hit the product, itself or any other component that is inside the workcell.

These criteria have to be jointly fulfilled and constitute the boundary conditions of the path planning problem. In order to identify a single solution for the path, an optimization process is implemented to minimize the time needed for completing the whole path. In order to make sure that the robots smoothly follow the path and do not permanently change its configuration, it usually pays off to split the whole part into different zones that are treated sequentially. Quite often, such zones can be geometrical primitives such as (parts of) cylinders, which enable the use of pre-defined strategies for path planning.

Path planning has been investigated for many camera-based inspection problems, such as completeness inspection [9], for automated 3D object reconstruction [10] or dimensional measurement using laser scanners [11]. For such camera-based processes, it is also known as "view planning problem". Usually, the length of the path or the number of views is chosen as an optimization criterion. In order to solve the associated optimization problem, algorithms that can deal with discrete search spaces, such as genetic algorithms [12] or particle swarm optimization, are regularly used. The basic problem structure is also known as the "art gallery problem" or the "watchman route problem".

Path planning is a computationally hard problem [13], so that even with a low number of waypoints finding a globally optimal path often is computationally infeasible. Moreover, a globally optimal path with focus on path length is not always desirable because other attributes such as the ability for dynamic replanning can also be important. Recently, large neighbourhood search [14] has shown remarkable results for path planning in constrained environments. Also, ant colony optimization [15] has been successfully used to solve large-scale route planning problems with hard constraints [16]. We will not go into further detail on how the path is planned and converted into a robot program. We refer the reader to the literature above and focus our discussion only on the machine vision aspects of the path planning.

At the heart of the path planning, algorithm is a process model that represents the image acquisition process. This process model answers the question, which area on the part can be inspected from a given viewpoint. It includes not only basic elements such as the optical properties (focal plane, field of view, etc.), but also more complex ones such as the range of object distances that are still in focus and that

deliver images of sufficient quality so that they are suitable for inspection. Some of these parameters can only be determined by experiments. The process model is highly specific for the particular image acquisition set-up that is used. In the following sections, we will provide two concrete examples of such process models.

The second major offline step is the calibration of the whole inspection system. Calibration provides information about the relative position and orientation of all the coordinate systems that are involved in the inspection task. There are multiple coordinate systems that we will discuss in the following paragraphs. The first one is the world coordinate system. This system is usually linked to some significant physical object in the workcell, whose position can be determined with sufficient accuracy. The world coordinate system is often used for simulations, where all objects in the robotic workcell are represented and need to be placed in 3D space relative to that coordinate system. From the world coordinate system, we proceed to the tool centre point of the robot. The tool centre point is linked to the world coordinate system via a kinematic chain across all the joints of the robot. This kinematic chain and the coordinate transformations associated with it are assumed to be known with sufficient accuracy. They are defined by the type of robot and are part of the specification of the robot. The position of the tool centre point relative to the world coordinate system is calibrated by making the robot point to a few particular, well-defined positions in the world coordinate system. Depending on the accuracy that is required, a pointing tool, such as a needle or stylus, may be used. The calibration procedure is often semi-automatic, where the needle is manually positioned and the transformations are calculated automatically. This functionality is typically a part of the robot's control software and does not need any further consideration when setting up a robotic inspection system. Assuming that the camera is mounted on the robot's gripper, a calibration is needed between the tool centre point and the camera coordinate system. This is called hand–eye calibration, and many different methods have so far been developed for this purpose [17–19]. Camera calibration involves two steps that may be executed separately or jointly: the intrinsic calibration and the extrinsic calibration. The intrinsic calibration is used to determine all parameters that belong to the optical system of the camera, such as the focal length, the centre of the optical axis, scaling and various parameters related to optical distortions. The extrinsic calibration determines the position of the camera, i.e. the centre of the optical axis, relative to other coordinate systems, in our case relative to the tool centre point. Intrinsic calibration and extrinsic calibration are done by placing a calibration object in front of the camera in different positions (or by moving the robot) and acquiring a series of images. Quite often, the calibration object includes a checkerboard pattern [20, 21], whose corners can be accurately localized in the images. The coordinates of these corners are assumed to be known and are then set in relation to the pixel coordinates of the camera, so that the intrinsic and extrinsic camera parameters can be determined.

The last coordinate system that we need to consider is the workpiece coordinate system. Usually, the 3D shape of the product to be inspected is defined in a 3D CAD model that uses a particular coordinate system that is linked with certain features on the product. In order to inspect the product, the inspection system needs to know

where the product is placed in space. Quite often, it is not needed to determine the position of each single product to be inspected, because manufacturing tolerances are sufficiently low so that the products can be considered as being identical for the purpose of inspection. Therefore, the standard solution is to have a mechanical fixture in which the product is placed and accurately aligned relative to the fixture. By having a set of fiducial marks on the fixture, it can be accurately positioned relative to the robot and the world coordinate system. In many cases, the robot's control unit also supports the definition of a product coordinate system, e.g. by again using a needle or stylus to identify the position of the marks relative to the robot coordinate system. Alternatively, the camera may be used to localize the marks.

Once we performed all these calibration steps and all the coordinate systems are determined to the required level of accuracy, we are able to map defects that were found in the image onto the product in 3D space. Starting from the pixel coordinates of the defect in the camera image, we use the intrinsic calibration and extrinsic calibration to transform the position into the tool coordinate system. By recording all the joint angles at the time the image was taken, we can use the kinematic chain of the robot to transform the position into the world coordinate system and with a final transformation into the product coordinate system. Actually, it is not as simple as that, because the camera only provides 2D information that represents a straight line in 3D space on which the defect is localized. This line then needs to be intersected with the 3D model of the product to determine the location of the defect on the product, but in any case, the sequence of coordinate transformations is the same as described before.

The final element that has to be preset before the automatic inspection can commence is the quality criterion that has to be applied. These criteria convert the list of defects into a single "accept"/"reject" classification. The criteria will be different for different areas on the part. For regions, such as sealing areas, tighter criteria will apply than for regions that do not fulfil a particular function. Initially, these criteria are specific in 3D, i.e. based on the CAD drawing of the part. In order to save computing time during online inspection, it is recommendable to pre-process the regions and compute back-projections of these regions into the single images taken from the different viewpoints. It is then much simpler and less time-consuming to identify those areas in the image that have to be inspected.

10.2.2 Online Components

After setting up the inspection parameters, the robot's path and all the coordinate systems, the robotic inspection system is ready for operation. During the execution of the inspection process, a set of distinct modules are operating in parallel that are described in the paragraphs below. The description will follow the logical processing steps from the image to the high-level decision-making as shown in Fig. 10.3 and will highlight the specific properties of these modules.
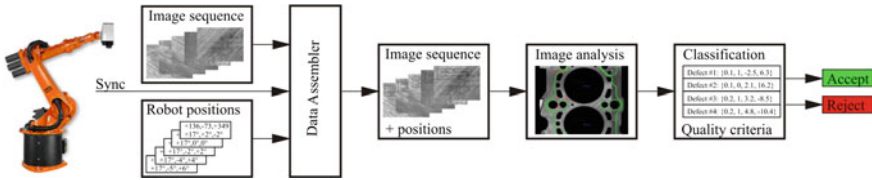
Fig. 10.3 Data flow during online processing

Processing starts with the images. Depending on the application, this may be just a single image or a set of images taken under different illumination. For an efficient implementation of the inspection process, we will avoid stop-and-go motion for each image and instead assume a continuous scanning process. This implies that all the images are taken at slightly different positions, which needs to be taken into account when the joint analysis of several images, e.g., with different illumination is needed for the application. For that kind of analysis, it is necessary to determine the exact location of a single, particular point on the part's surface in the set of images. Due to the continuous motion, this point will appear in different locations in the images. To facilitate image registration, accurate position information is required that usually comes from the robot's control unit. A common approach is to synchronize image acquisition and the robot's motion either electrically, where the camera provides an electronic trigger signal whenever an image is acquired, or by software, where time stamps are used to perform the synchronization. Recently, communication protocols such as the precision time protocol (PTP) have been developed to solve this problem. The process of merging image data and robot position information is done by a module that we would like to call data assembler. The task of this module is to create a sequence of position-referenced images, so that each image also contains information about the joint angles of the robot at the time of acquisition. The cycle times of the camera (e.g. 200 frames per second) may be different from the cycle times of the position read-out module (e.g. 10 ms), which may require interpolation between time instances.

The sequence of position-referenced images is then fed into an image processing module. The exact processing that takes place totally depends on the application. Examples of such processing will be described in the following sections. Common to most of the applications is the need to process comparably large amounts of data. This is best done by a pipeline-based architecture, where the single processing steps are done in parallel for multiple images taken at different times. This processing pipeline will ensure that the amount of data can be processed in time; however, it has to be noted that there will always be a delay between the time the image is taken and the time by which the processing result is available. However, in robotic inspection systems, where we need to wait until the inspection of the whole part is finished, this will not be a problem. Typically, we expect delays of about 1 s, while inspection of the whole part takes 30–60 s.

The sequence of processing steps follows a typical scheme that is widely used in surface inspection tasks. The first step is the image segmentation. This step is highly

application specific and will look totally different, depending on the surface texture, the reflection properties and on the physical processes (e.g. in thermography) that are guiding the image acquisition. Goal of the segmentation is to extract a set of image patches that show "interesting" areas. We call these "areas of interest" or "events", because the fact that it is detected by image segmentation does not necessarily imply that it is a real defect. To determine whether there is a real defect, the image patches are converted into feature vectors. Again, the features may be application dependent; however, at this processing level, quite generic features exist that are useful for a very wide range of inspection tasks. The feature vectors are then fed into a classification system that determines whether there is a defect, and if so, the particular type of defect is shown. Features used in this classification may include the grey-level distribution inside of the defect (texture) or depth information coming from photometric stereo calculations. Sometimes, calculation of these features can be optimized to improve classification accuracy [22].The main challenge is to distinguish the real defects from the many other things that may appear on the surface.

The final step after classifying the different types of defects is an accept/reject decision for the whole part. There are two substantially different cases. In the first—more seldom—case, the decision is based on the subjective judgment of an expert doing quality control. This happens, when aesthetics have to be judged or other properties that are difficult to assess in an objective manner. In order to convert this expert's knowledge into an automatic decision, machine learning methods are used that extract decision rules from a set of samples [23]. In the second case, the decision is based on quality control criteria that are documented in a quality management system. This criteria are written down in the form of rules including specific thresholds that determine limits for properties such as the maximum size of a defect, the minimal allowed distance between defects, the maximum number of defects and other rules that deal, e.g., with clusters of defects. Quite often, however, we observe differences between the actual manual inspection process and the quality criteria. Human experts who performed the inspection often gather specific knowledge about the production processes and adjust their decision-making based on their observations [24]. These adjustments are typically not documented in the quality management system. In order to make sure that the inspection robot best reproduces the actual (human) inspection process, statistical analysis of samples and machine learning will be applied here as well.

10.3 Carbon Fibre Part Inspection

Composite materials are nowadays gaining more and more importance in many branches of industry. Classically, composite materials were limited to very special applications such as aerospace or high-end sports equipment. However, the use of these materials is getting more and more popular for many different applications. The widespread use in a large spectrum of products increases the need for

production processes with a high degree of automation and reduced costs. In this context, automated inspection and process control based on machine vision are very fruitful technologies.

In the following, we focus on surface inspection of carbon fibre and glass fibre-reinforced plastics. These materials exhibit rather complex optical properties as they are shiny and provide either too little or too high contrast. This results in too dark or too bright patches in the images and prevents further processing. With respect to the design of a machine vision system, the following properties of fibre-reinforced plastics are thus relevant:

- The mirror-like specular reflection (carbon fibre and glass fibre behave like tiny cylindrical mirrors) puts several challenges on a reliable machine vision system.
- Raw textile fibre materials typically exhibit a repeating pattern (e.g. woven material, sewing yarns). Such repeating patterns make it difficult to stitch images by means of computer vision methods [25, 26].

Regarding surface errors that need to be detected by a surface inspection system, two relevant scenarios are identified:

- Inline inspection for automated post-treatment.
- Inspection for final quality control and good/bad decision.

A crucial property of fibre-reinforced plastics is the orientation of fibres within the material. Fibre orientation matters because of two different reasons. First, mechanical properties depend strongly on the orientation of the fibres. A structural carbon fibre part can only deal with strong forces that act upon it, if forces and carbon fibres are well aligned. This is often designed using finite element calculations, and much effort is spent on minimizing the weight for given mechanical properties and loads. Second, for those parts that are visible in the final product, fibre orientation matters because of aesthetic reasons. Typically, fibres should be aligned with the edges of the part, and distortions of the woven material are not allowed.

In the remainder of this section, we describe a robotic surface inspection system that inspects fibre angles on the surface of 3D parts with moderate geometric complexity. The system is able to determine the fibre angles on the surface of carbon fibre or glass fibre-reinforced plastics, fully covering the part. Of course, there are some physical limitations, e.g., in tight corners of the parts, in areas that cannot be reached by the robot and in places where collisions would occur.

We focus on the surfaces of so-called pre-forms. In the production process of fibre-reinforced plastics (FRP), pre-forms represent an intermediate step. Basically, a pre-form consists of fibre textiles which are already put roughly into the shape of the final part. In subsequent production steps, resin is injected into the part or the pre-impregnated pre-form is heated. In both cases, the liquid resin tightly fills all the space between the fibres. After curing, the part keeps its shape on its own. A critical point is that pre-forms are easily deformable. In this state of the production process, the fibres are not kept in place by the surrounding plastic matrix. As a result, the actual shape of the part does not perfectly match the CAD data, which requires us to first obtain sufficiently accurate 3D data. Hence, we propose a separate geometry capturing

step using a laser range scanner or low-cost alternatives such as RGB-D sensors which are gaining more and more popularity for 3D object reconstruction [27].

For the inspection of pre-forms, it is in general more suitable to mount the camera on the robot and move it over the part. Keeping the camera fixed and moving the pre-form part is not an option as the pre-form would deform and could even be damaged. The presented system supports two modes: stop-and-go inspection, where the robot is stopped for capturing a set of images and continuous capturing of images. While continuous capturing of images is the far more interesting approach, we will also discuss aspects of stop-and-go inspection.

10.3.1 Image Acquisition Set-up

There exist different approaches to perform automated inspection of carbon fibre-reinforced plastics. We shall ignore general non-destructive testing methods such as computed tomography, ultrasound and the like and focus on machine vision. One possible approach to inspection using machine vision is to suppress specular reflections of carbon fibre and glass fibre. This may be tackled with diffuse illumination of the inspected surfaces [28]. The resulting image ideally does not show any specular reflections and is homogeneous, occasionally with low contrast. Texture analysis methods can then be used to calculate fibre angles or perform defect detection. However, this method proved to be somewhat sensitive to changes in the illumination and to the properties of the material that is inspected. Also there are difficulties when inspecting clear-coated parts. We are thus following a different strategy for the described fibre angle measurement system. Instead of suppressing the specular nature of fibres, we use a reflection model that exploits the specular reflectance of the fibres [29, 30]. Based on a set of images that are taken under different illumination conditions, the reflection model makes it possible to calculate fibre orientations. Basically, a special form of photometric stereo [31, 32] is applied.

Carbon fibre and glass fibre are modelled as very thin cylindrical mirrors. A light ray that hits a small cylindrical mirror is reflected in the shape of a cone. Typically, carbon fibres have a diameter of approximately 5–9 μ, while camera-based surface inspection systems cover a surface region of at least 20×20 μm^2 per pixel. Hence, the corresponding light rays have a sufficiently large diameter, so that the cylinder may be considered as being infinitely small and the cone-shaped reflection model remains valid.

Given a fixed light position, the exact shape and orientation of the reflection cone depend on the orientation of the reflecting fibre. Inversely, given the parameters of the reflection cone, it is possible to determine the orientation of the fibre in 3D space. By capturing images illuminated from different directions, it is possible to determine the reflection cone parameters and, hence, also the fibre orientation on the surface.

For the following considerations, we concentrate on the inverse direction of light. We follow the line of sight from the camera's optical centre (denoted c in Fig. 10.4) to the intersection with a fibre (denoted O). According to the
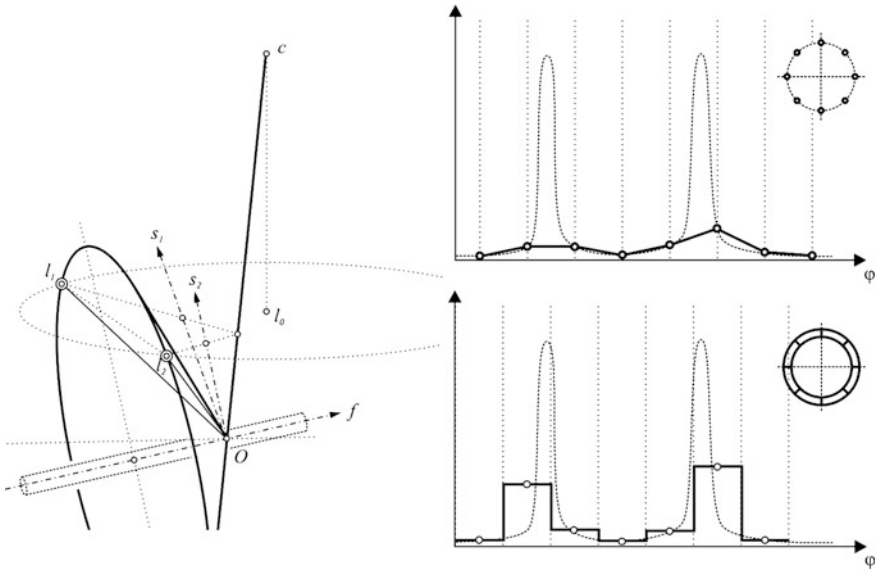
Fig. 10.4 Fibre reflection analysis (from [29])

considerations above, this ray is also "reflected" in the shape of a cone. Light sources are located on a circle with centre l_0. The intersections of the light circle and the reflection cone are marked as positions l1 and l2. In an ideal model, only light from these two positions on the circle is reflected into the camera.

By capturing many images with different light sources switched on, the intersection of the reflection cone of the line of sight with the circle of illuminations is calculated. This is done by investigating the variation of brightness of pixels corresponding to a single point on the surface. In case of point light sources, theoretically, an infinite number of light sources on the circle are required in order to exactly hit the intersection points. The diagram on the right top of Fig. 10.4 shows the variation of grey values for corresponding pixels with point light sources. The dotted line shows the intensity of reflections for an infinite dense number of light sources. The continuous line shows the intensities for a limited number of 8 light sources. The peaks of the dotted line indicate the intersection points of light circle and reflection cone. Knowing these intersection points, it is easy to calculate two normal vectors s_1 and s_2 of the fibre that points into direction f. Finally, the cross-product of s_1 and s_2 is equal to f.

A dense distribution of light sources over the light ring is theoretically necessary. In order to solve this problem, broader segments of light sources are used. A set of 8 light segments are sufficient to reconstruct the distribution of grey values (Fig. 10.4, right bottom).

The approach that has been described above works well for static inspection, where the camera remains in the same place until all the images under different illumination have been acquired. If the camera is moving during the acquisition
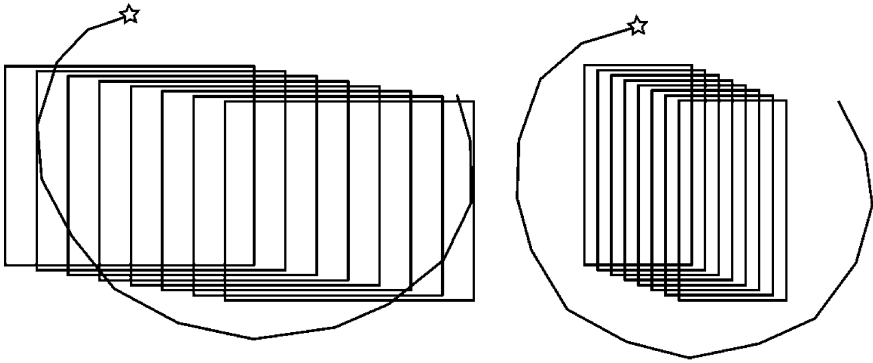
Fig. 10.5 Image format for continuous capturing during motion—large format (*Left*) versus slim format (*Right*)

process, accurate registration between the images is needed. This raises a few open questions that will be addressed in the next paragraphs. First, we need to investigate the optimal size of images that should be captured. Many industrial cameras offer the capability of acquiring only a subset of all image pixels at the benefit of higher frame rates. Here, we consider two possible strategies: large (slow) versus slim (fast) image format. When capturing large images during motion, the individual raw images are captured at quite large distances (Fig. 10.5, left). Using a slim image format, images are captured at relatively short distances (Fig. 10.5, right). Note that the overlap between consecutive images of about 90 % is identical in both cases. Three issues need to be considered regarding the image format:

1. Inaccuracy in registration: in order to perform the previously described fibre reflection analysis, the very same surface point needs to be identified in a set of sequentially captured images. For a typical set-up with 8 raw images for fibre reflection analysis, a point needs to be visible in all 8 raw images. If the captured images are rather large, the distance between two successive images has to be large as well in order to minimize the time needed for inspection and avoid unnecessary overlaps. If the images are captured at large intervals, the large distance of two successive capture positions makes it difficult to establish accurate pixel-to-pixel correspondences.
2. Cycloid movement of light positions: in a static set-up, the surface is illuminated with light coming from different positions along a circle. If the camera is moving together with the light sources, the circle becomes a cycloid. If the temporal and spatial distance between two consecutive images is large, then this cycloid shape has to be considered in the analysis.
3. Curvature of surface transversal to motion: unlike the two previous points, this one concerns the size of the image in the direction that is orthogonal to the scanning direction. If the surface is strongly curved in this direction, the usable part of the captured image will be small.

   Given these considerations, the general preference is to have a small image height in direction of the scanning motion and reasonably wide images orthogonal to the scanning direction. This simplifies image registration and optimizes the speed. Typical frame rates go up to 1000 images per second.

   In order to assign the correct sensor position to the captured images, it is necessary to synchronize image acquisition with robot positions. We propose the use of time stamps to achieve this synchronization. An integrated clock is a standard feature in many modern GigE industrial cameras. This integrated clock is used to assign a time stamp to each image. Also, robot controllers usually contain a clock which we use to assign time stamps to continuously record robot positions. In fixed time intervals, the clocks of camera and robot controller are synchronized by calculation of the time offset. Clock synchronization makes it possible to correctly assign robot positions to the captured images. We are using a software implementation of this synchronization although hardware solutions (e.g. PTP hardware implementation) are also available and provide even higher accuracy.

10.3.2   Process Model

The basis for the process model of image formation is the fibre reflection analysis described in the previous section. The critical question which the process model has to answer is as follows: for which regions does the fibre reflection analysis work, given the position of the sensor relative to the surface. Of course, this includes the question which regions are basically visible by the camera system. This covers aspects such as the camera's field of view and in-focus depth. However, additional aspects need to be considered as well: in order to make fibre reflection analysis possible, it is required that the investigated fibres' orientations are within certain limits. Figure 10.6 shows typical results of fibre reflection analysis. In the leftmost image, the fibre angles are colour-coded, with red being horizontal (left-to-right) fibres and blue being vertical (top-to-bottom) fibres. The third image from left shows the same material wrapped around a cylinder. It is clearly visible that the analysis fails in regions where the surface is tilted with respect to the imaging plane. In case the reflection analysis fails, more or less random fibre angles are calculated, which corresponds to the noisy image regions.
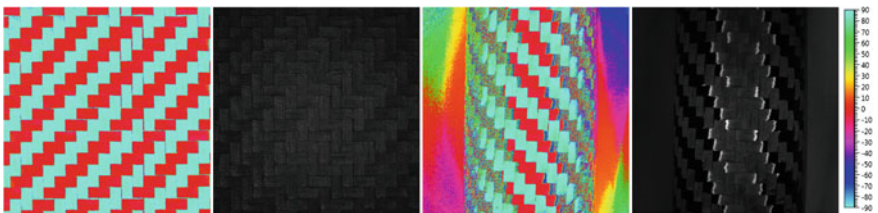
Fig. 10.6   Fibre reflection analysis—flat surface (*Left*) versus cylindrical surface (*Right*)
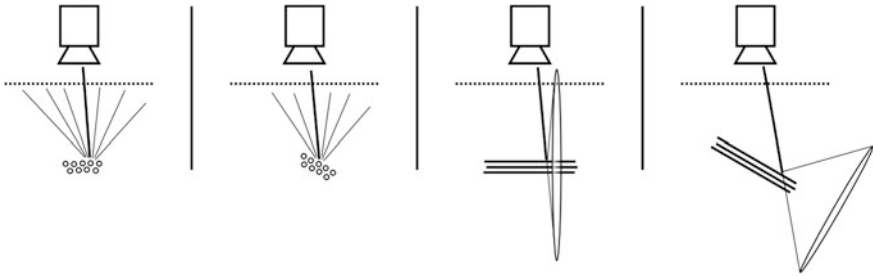
Fig. 10.7 Illustration of fibre reflection for different combinations of in-plane and out-of-plane orientation. Fibre reflection analysis fails if no light from the LED ring reaches the camera (right most illustration)

For the cylindrical shape we see that not only the overall surface geometry influences the success of the analysis, but also the fibre orientation is relevant. Whereas for the horizontal (red) fibres the analysis fails already at low deflections ($\sim 12°$ out of the imaging plane), correct calculations can still be done for (blue) fibres that run from top to bottom at angles of up to 40°. An explanation for this is illustrated in Fig. 10.7.

Based on this process model, we can describe at which fibre angles the analysis fails. However, at the point in time when offline path planning is made, the actual fibre orientations are not yet known. Even though (in-plane) fibre orientations are documented together with the CAD data ("ply book"), the fibre angles may deviate from these ideal orientations. The surface geometry that we know a priori somehow restricts fibre orientations, as it is reasonable to assume that fibres are almost parallel to the surface and out-of-plane angles are close to zero. However, for some materials (e.g. woven textiles) or defects that occur, the critical out-of-plane angle may be higher than expected and subsequently may lead to failure of fibre reflection analysis.

For some applications, it may be possible to safely estimate the range of expected out-of-plane angles. In this case, this range of angles is used in the offline path planning and failure of fibre reflection analysis is excluded. The drawback is that the estimated range of out-of-plane angles may be larger than actual out-of-plane angles. If this is the case, path planning may lead to a path that is much more complex than necessary. In order to avoid unnecessarily complex paths, dynamic replanning of inspection positions may be implemented. As soon as a failure of fibre reflection analysis is detected, the relevant surface region is captured from additional varying positions until a valid fibre angle can be calculated. The critical point is that dynamic replanning should not occur too often as it may introduce a significant increase of scanning time. Furthermore, it is not straightforward to implement automatic detection of failure of fibre angle calculation. One possible approach is to assess the noise level of calculated fibre orientations. If the noise level is too high for some image region, a failure is assumed. In general, the strategy for dynamic replanning and a reasonable range for expected out-of-plane fibre angles need to be adapted to the respective application. Relevant parameters are mainly the material used and the types of possible defects.
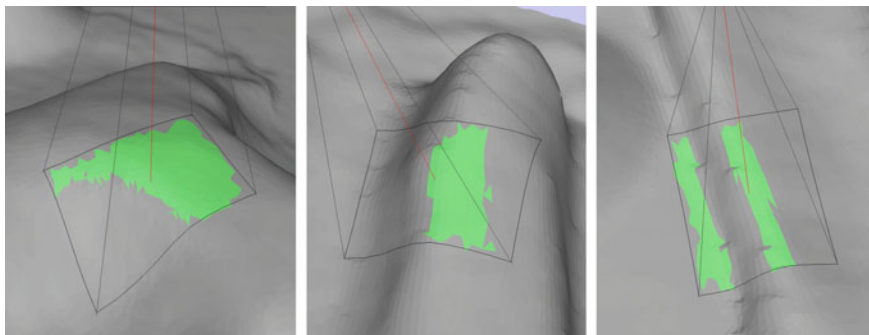
Fig. 10.8 The process model is used to calculate valid regions in the field of view. The *black* boundary marks the visible region of the camera. Valid regions (with respect to the process model) are c*oloured green* (colour figure online)

Figure 10.8 shows different examples for results provided by the process model. Visible regions as seen from the inspection system are marked with black lines. The actually valid regions for which fibre reflection analysis is expected to work are restricted to much smaller regions coloured green.

10.3.3   *Image Analysis*

Once a full scan of a part is finished, fibre reflection analysis is performed using the captured image data from the 3D surface. Two approaches are possible:

1. Starting from the 3D CAD model, we identify the point for which we want to calculate the fibre orientation. For this surface point, the set of images needs to be determined that shows the point under different illuminations. By projecting the point into the images, we extract the single grey values and calculate the fibre orientation. This requires that the projection of each surface point in several images has to be calculated as it is not self-evident in which images the point will be visible. Nevertheless, this approach makes it easy to calculate fibre angles on a regular grid mapped to the 3D surface.
2. Fibre orientation can also be calculated in the 2D domain. Because the images stem from 3D surfaces, a mapping of surface points to 2D points is introduced and a transformation between consecutive images in 2D is calculated. For almost flat surfaces, this may be approximated with an affine transformation. For more complex surface geometries, also more general distortions need to be considered. The fact that multiple pixels (whole images) are transformed and analysed at once makes this approach very efficient. Once the fibre angles are calculated in 2D, they are projected back onto the 3D surface in order to obtain the fibre orientations on the inspected surface. With this approach, a dense mapping can be more easily achieved; however, the density on the 3D surface is not regular and will vary depending on the geometry.

Independent of the exact strategy for fibre angle calculation for individual surface points, the calibration accuracy is very critical. This is the case because an exact pixel-to-pixel mapping between individual raw images is desirable. For the described inspection system, a single pixel maps to a surface region with a diameter of approximately 20–80 μm. To establish an accuracy of positioning for pixel-to-pixel mapping at this scale is clearly a challenge. Of course, calibration of the whole inspection system including an accurate localization of the scanned part has to be done. In order to further increase the accuracy of image registration, image stitching methods that work on 3D surfaces [33] may be considered. The idea is to define an objective function that describes how well individual texture patches overlap. The objective function is optimized subject to parameters that describe the location of images projected onto the surface.

Most of the existing stitching methods aim at texturing of objects under diffuse illumination. In this context, it is comparatively easy to set up a reasonable objective function, e.g. based on a pixel-wise comparison of grey values. For fibre reflection analysis, the images are acquired under different illumination and images taken from the same spot may look very different. A pixel-based comparison of grey values is thus not likely to succeed. Instead, the noise level in the calculated fibre orientations may be used as an objective function for image alignment. General frequency measures of combined images such as those used in autofocus for cameras are also applicable.

10.3.4 Prototype Implementation

A prototype of a CFRP inspection system was implemented using a TX90L Stäubli industrial robot. The sensor system is based on a Genie TS 2500M industrial Gigabit Ethernet camera and a high-power LED light ring. A dedicated microcontroller triggers LED light sources and image acquisition of the camera. Different image formats, LED lighting sequences and other settings of the image acquisition system are supported and can be adapted to the respective application.

Figure 10.9 shows a picture of the sensor system together with a 3D visualization of the robot, the sensor viewing pyramid and the CAD model of the part. In the visualization, fibre orientations for a part of the CAD model are colour-coded and back-projected onto the CAD surface. Figure 10.10 shows the resulting fibre angles (colour-coded) for a region on a curved part.

10.4 Thermographic Crack Detection

Zero failure production is gaining more and more importance in industrial production processes for two main reasons: first in respect to reduce production cost due to minimizing waste material and second to ensure highest product quality for
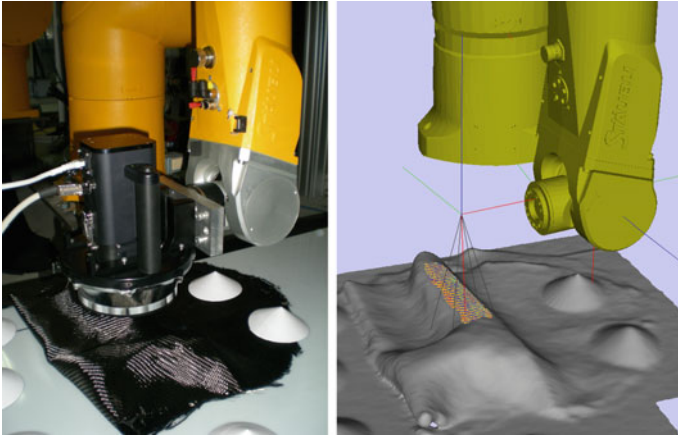
Fig. 10.9 Prototype implementation: fibre orientation sensor scanning the surface of a part—picture (*Left*) and visualization (*Right*)
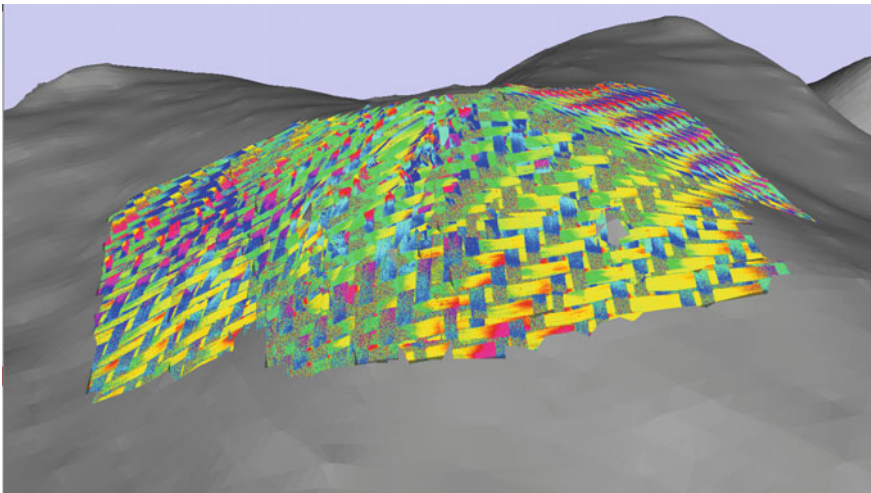
Fig. 10.10 Fibre angles mapped onto the surface of 3D-shaped part

the whole product life cycle time in order to avoid expensive callback actions. In particular, components working under strong mechanical and thermal stress need to be checked very carefully, since even small defects can affect performance and reliability in a negative way. Cracks are a common source of failure in metallic parts and thus require much effort for reliable crack checking.

The most common current procedure for crack detection for metallic parts is a process that dates back to the 1920s and is called "magnetic particle inspection" (MPI). This method is infamous in industry, because it is a cumbersome, dirty

process that is often done manually even in otherwise fully automatic production lines. The component to be tested is temporarily magnetized before applying a suspension of finely divided coloured or fluorescent magnetic particles. Cracks or inclusions disrupt the magnetic field and cause the magnetic particles to show up the crack. Using UV light under low-light conditions, the fluorescent particles in the suspension are activated to increase the visibility of the cracks. Afterwards, the component needs to be demagnetized. Magnetic particle inspection for complex parts is a manual process with well-known uncertainties in inspection performance. In combination with the necessity for 100 % inspection, an automated solution for crack detection is inevitable.

In the following, we focus on a non-destructive method based on infrared thermography which can be used for fully automated crack detection. Infrared thermography for non-destructive testing aims at the detection of (sub-) surface features, i.e. cracks or material anomalies, owing to temperature differences observed on the investigated surface during monitoring by an infrared or thermal camera [34]. An infrared camera detects and records the radiation emitted (electromagnetic radiation) by a material under investigation and converts this into a thermal image [35].

In general, infrared thermography is specified in two categories: passive and active. Passive thermography is used to investigate objects which are at a different temperature than ambient temperature. Most common applications for passive thermography are for thermal efficiency survey for buildings and infrastructure, predictive maintenance, medicine, fire detection and in non-destructive testing [36]. Active thermography uses an external heat source (i.e. flash lamps or laser) in order to induce a spatially and timely well-defined heat pulse to the component to be inspected causing a transient phase as long as heat distributes into the sound areas. This heat flux will be disturbed by voids such as surface or subsurface cracks. Fig. 10.11 illustrates the typical heat dissipation caused by a spatially limited heating. For homogenous materials, the heat dissipates uniform in all directions. Cracks disrupt the heat dissipation and cause non-uniform heat dissipation which can be detected by a thermal camera. Figure 10.11 illustrates the expectable heat dissipation caused by a point heat source and disruption by a surface near crack (right) and a crack at deeper location.
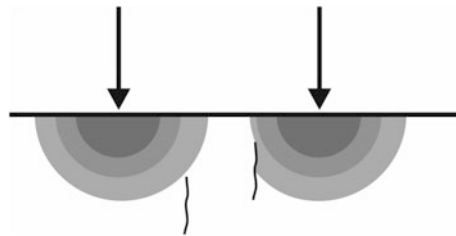
Fig. 10.11 Detection of subsurface cracks. The crack depth must be much smaller than its distance to the heat source

In the following, we focus on crack detection for metallic forged parts, more precisely for crankshafts. Crankshafts are available at a high number of different shapes and materials. They are parts of a high geometric complexity and possess different surface qualities. Typically, surface qualities of forged parts vary in the rang from rough unprocessed surfaces with typical emission coefficients of about 0,8 up to shiny processed surfaces with typical emission coefficient of 0,2. Basically, crack characteristics for the considered parts vary in a wide range in respect of position, orientation, size and depth. Cracks may also be closed at the product surface.

In the remainder of this section, we describe a robotic crack detection system which is able to inspect metallic forged parts of complex geometries using heat flux thermography. This system is able to perform a 100 % crack inspection of the sample parts. Of course, there are some physical limitations, e.g., in small gaps or areas that cannot be reached by camera and laser or regions where collisions may occur.

10.4.1 Image Acquisition Set-up

There exist several non-destructive testing methods to perform automated crack detection, for example eddy current, ultra sonic or X-ray. We focus on thermographic crack detection. One possible approach is to use passive thermography. In case of crack detection, passive thermography works only for surface-breaking cracks and materials with low-emission coefficients and requires the component to be tested to be at a significant higher temperature than the ambient. Caused by the different emission coefficients between material (low-emission coefficient) and crack (high emission coefficient), the crack occurs brighter than the surrounding material and can therefore easily be detected within the thermal image. Figure 10.12 shows two sample images acquired by using passive thermography methods. The left image shows a key with fatigue crack. The right image shows the front side of a gearwheel with a crack on one of its tooth.
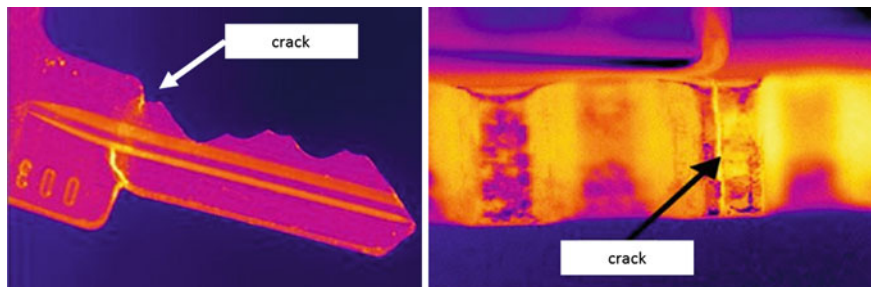
Fig. 10.12 Sample images fro crack detection using passive thermography. The images show a key with fatigue crack (*Left*) and a gear-wheel with a crack on the front side of one tooth (*Right*)

However, this method proved to be pretty sensitive to surface qualities and does not perform for cracks which are closed or covered at the surface. We are thus focusing on active or transient thermography. In the case of active thermography, an external stimulus is necessary to induce relevant thermal contrast.

Inductive heating is a widely used and effective excitation method for crack detection of ferritic metal parts, e.g. for inline inspection of steel billets [37]. The technique uses induced eddy current to heat the material. Hence, inductive excitation is limited to be used for sample parts made from electric conductive materials and to detect cracks with known predominant crack directions. Since crack characteristics for crankshafts do not follow any predominant direction, we are thus following a more flexible strategy for heat induction by using laser spot heating.

Starting with a test part at thermal equilibrium, laser excitation induces a spatially and temporally well-defined heat impulse to the specimen. A spatially limited area (affected by the laser sport) will be heated compared to the sound material. This local temperature difference between spot position and sound area acts a motor for the heat flux. Heat starts to dissipate into the sound material. Heat flows until temporal equilibrium is reached. For homogenous materials, the heat flux is uniform and will be disrupted by any crack or void. The time period of the transient phase depends on the amount of induced energy (local temperature difference) and the heat conductivity of the material. Basically, metals are good heat conductors so the transient phase is temporal and spatially limited to a very short time period after excitation and area near the excitation position. By moving the laser spot, larger areas and hence whole test sample can be inspected. Figure 10.13 shows thermal images of a ring-shaped object with crack during laser excitation at different points of time (t_1 until t_6). The laser spot moves from the left (t_1) to the right side (t_6). The crack position is marked by a red dotted line. It can be seen that in t1, the laser spot is far away from the crack position, and the heat flux is not disrupted by the crack. The closer the spot position and crack position are, the bigger the distortion of the heat flux until the laser spot passes the crack position.
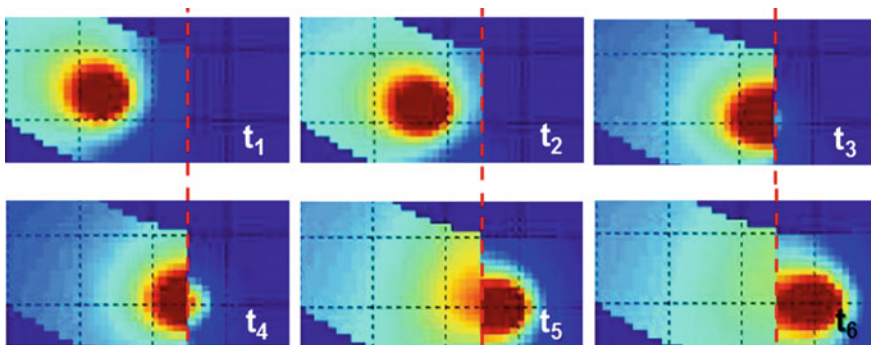
Fig. 10.13 Thermal images of a ring-shaped object with a laser spot moving across captured at different points in time. Laser moves from the left (t_1) to the right (t_6), crack position is marked by the *red dashed* line (colour figure online)

The approach that has been described above works well for static inspection where the camera and the object stay static until the image sequence is fully captured. If the camera or the object is moving during the acquisition phase, which is necessary in order to cover larger or complex-shaped objects, an accurate registration between the images is required. Image registration can be performed by using tracking methods. Besides sufficient overlapping areas between consecutive images, this method requires sufficient number of traceable features within every single image. Cracks or other voids can be used for feature tracking. However, this method fails in case of the lack of sufficient image features. In case of thermal imaging, this may be the case if there are no cracks or visible. A more general approach for image registration deals with the acquisition of additional time and position information for each single image. Using this information, the offset between two consecutive images can be exactly determined.

10.4.2 Process Model

The process model for a robot-based thermographic crack detection needs to deal with the heat dissipation, and therefore, material and heat source properties are the main input values. The critical question which the process model has to answer is as follows: for which areas does the crack detection work, given the position of laser, camera and test sample. Of course, this includes the question which areas are visible for the thermal camera and are at the same time in a direct light path of the laser. This covers optical aspects such as the camera's field of view and focal depth. However, additional thermal aspects such as thermal conductivity, emission coefficient, laser power, velocity, and distribution and geometry of the test sample need to be considered.

For the case of an object with flat surface and homogeneous material placed at an ideal angle under the camera and laser, the inspectable area will have a ring-shaped form as shown in Fig. 10.14. At the centre of the ring, there is the laser spot which results in a bright, white region in the image, where the pixels of the thermocamera are fully saturated and no analysis is possible. As the heat dissipates isotropically,
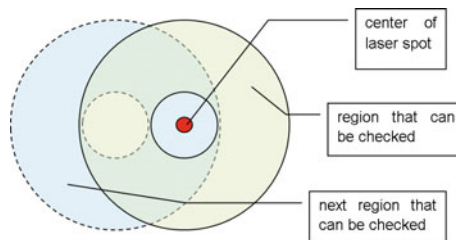
Fig. 10.14 Basic concept for the area that can be inspected for laser-induced heat flux evaluation (from [36])
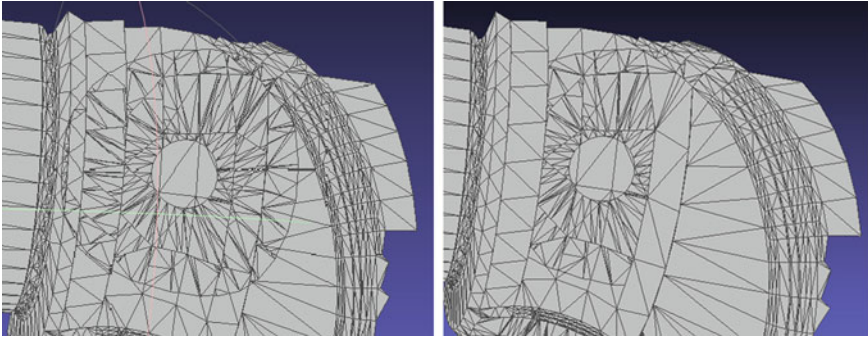
Fig. 10.15 Projection of the process model results onto the 3D shape of the parts surface. *Left* projection on parts surface. *Right* reduction to feasible area

the signal becomes weaker until no more contrast is achieved. This area defines the outer edge of the ring-shaped region.

During the inspection process, the ring-shaped inspection area is moving across the surface leaving a trail where the part has been inspected. Cracks will become visible as a temperature gradient that is caused by interruption of the heat flux. It should be noted that the sensitivity of the detection depends on the orientation of the crack relative to the laser spot. If the crack is propagating radially from the laser spot, it will not be visible as there is no heat flow across the crack.

In the more realistic case of a non-flat part, the situation becomes significantly more complex. The model has to consider that laser and camera are not placed in an ideal position relative to the part's surfaces and that the heat propagates in a non-flat area. An approximation of the area that can be checked may be obtained by projecting the ring-shaped region onto the part's 3D surface. Additionally, self-occlusions of the part have to be considered as well as areas of high curvature, where the above-mentioned approximation is invalid. Those areas have to be excluded from the region that can be checked. Figure 10.15 shows one projection.

10.4.3 Image Analysis

Once the thermal image sequences are acquired, crack detection analysis is performed using the captured thermal images. Basically, there are two different approaches for crack analysis: the first deals with methods applied on single thermal images, and the second approach takes the temporal aspects of heat flux phenomena into account. Typically, the crack analysis is split into several processing step:

- The first step or pre-processing step deals with non-uniform correction (NUC), bad pixel replacement and calculation of temperature values if required.
- Second, the image analysis step works for single thermal images and deals with image registration, definition of area of interest (AOI), exclusion of the laser spot centre, gradient calculation and (crack) pixel classification.

- Temporal analysis step deals with the accumulation of crack pixels and morphological processing, e.g. closing algorithm, segmentation and classification.
- The final analysis step covers the transfer of resulting data into real-world coordinates and displays the results in a 3D model of the testing sample.

Above-mentioned processing steps are typically applied using intensity ("grey") values rather than real temperature values. Since "grey" values are represented in 16-bit integer values, while temperature values are typically represented as floating point values, this approach significantly decreases processing time and memory requirements compared with the usage of temperature values.

When using infrared cameras before applying any image analysis or classification methods, prior data pre-processing is required. The pre-processing step is basically equal for all thermographic applications. Caused by the manufacturing process of thermal focal plane arrays (FPA), each pixel has its own gain and offset values. The non-uniform correction (NUC) aims at the determination of these individual gain and offset values. This can be done from two thermal images, one captured at low temperature and the second captured at high temperature. The calculated per pixel offset and gain values are then used as input values for NUC applied for all subsequent captured images. For some of the sensor pixels, typically less than 1 %, the NUC does not work. These so-called bad pixels need to be identified and replaced by an interpolation between neighbouring values. This process is called bad pixel replacement (BRP). Different kinds of thermal cameras require different intervals for renewal of NUC and BRP. For bolometer cameras, the interval for recalculation of NUC and BPR is typically some minutes, and for quantum detector cameras, e.g. with an InSb-Chip, the required NUC renewal interval is typically some hours. For some applications, the temperature values are required for subsequent analysis. Calculation of temperature values can be done by using either a parametric model or calibration data. Temperature calculation is simple for surfaces with emission coefficients of 1, which is only true for blackbodies. For objects with lower emission coefficients, like metals, the actual emission coefficient needs to be determined or measured in order to calculate proper temperature values. Once the pre-processing step has been finished, the heat flux analysis can be performed.

The image analysis step for laser-induced crack detection deals with the analysis for heat flux discontinuities or interruptions. The area with best sensitivity is described by the highest heat flux values. Figure 10.16 shows a typical temperature distribution after laser excitation (right) and the corresponding heat flux values (left). Sensitivity for crack detection is high for high heat flux values and low for areas with low heat flux values. Since the heat flux is weak for the laser spot centre and for remote areas, the crack analysis focuses on a ring-shaped area with the laser spot as its centre position.

Identification of the evaluation area starts with the detection of the laser spot position. This can be performed either by the usage of coordinate transformation methods or simply by hot spot localization. Once the spot position and its size have been identified, the surrounding area can be defined as AOI and further processed
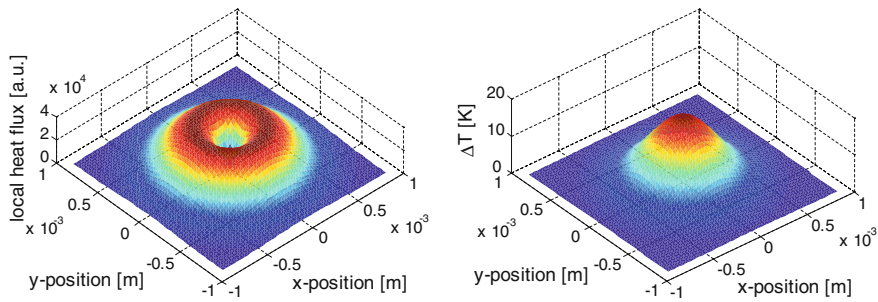
Fig. 10.16 Local heat flux (*Left*) and temperature distribution (*Right*) after laser spot heating

for crack analysis [38]. For crack analysis, several methods that are gradient based are feasible. As described in [38] radial gradient, tangential gradient and edge-based crack detection methods are used for heat flux analysis. Besides gradient amplitude, gradient direction can be used for discrimination between cracks and signal distortions caused by effects not wanted to cause a crack signal.

The temporal heat flux evaluation takes into account that signals caused by real heat flux are strictly time dependent as described by heat conduction. Signals caused by distortions like, e.g., reflections may not show this expected temporal behaviour. Temporal evaluation therefore supports noise suppression and more accurate crack detection. Furthermore, all results gained from single image analysis step are accumulated into a joint result image. Since all AOIs are defined in image coordinates, the results need to be transformed into real-world coordinates prior to result accumulation. Since cracks are visible in more than one AOI, typically within 10 consecutive AOIs, real cracks are amplified by adding the crack signals for each single image analysis to the result image, while artefacts occurring in one single image will be diminished. This leads to an optimized dynamic range between real crack signals and artefacts caused thermal signals.

The so-calculated resulting crack image shows single crack points coming from the results of the single image analysis. In the majority of cases, these crack points can be grouped into some kind of line shapes. For some circumstances, the crack line shape may be interrupted, maybe caused by signal noise or physical properties of the crack. Cracks are of a natural form with irregularities of shape, size, width and depth. For example, for areas with fewer crack width, the resulting image may show some interruptions although the crack physically is not interrupted but show varying widths. Using a morphologic closing algorithm, such gaps can be closed, while single defect pixels are eliminated. Crack length measurement and comparison with pre-defined minimum crack length values lead to an overall pass or fail decision. Figure 10.17 illustrates a sample of gap closing for line-orientated crack pixels.

The final processing step transforms the resulting image into real-world coordinates of the test sample. This transformation results in a 3D shape of the test sample displaying location and geometry of all detected cracks.
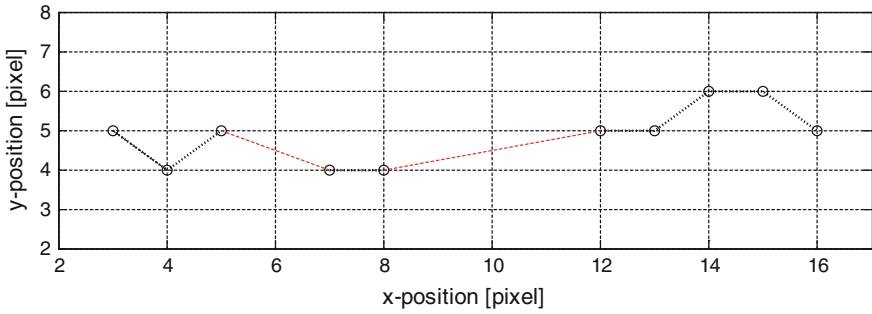
Fig. 10.17 Crack pixels (points) with *black* connection when line segment is closed within result image and *red* connection when points are missing (colour figure online)

10.4.4 Prototype Implementation

A prototype for thermographic crack detection was set up using a M710-iC Fanuc industrial robot. The robot was selected to carry a load of 50 kg so that it is suitable for the crankshaft. A water-cooled 100 W CO_2 laser is used as heating source. With respect to the available image formats, the camera type ImageIR® 8300 with a pixel number of 640×512 was chosen. This detector size provides an adequate field of view to cover the complete test sample surface with a certain geometrical resolution in an acceptable scanning time.

Figure 10.18 shows a picture of the prototype implementation of the thermographic crack detection system for crankshafts. The prototype system runs at an
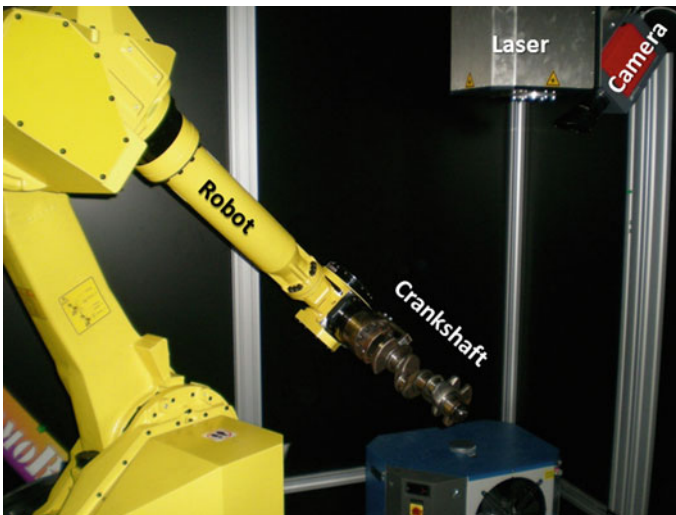
Fig. 10.18 Prototype implementation: thermographic crack detection for a crankshaft. The picture shows the robot holding the crankshaft, the thermal camera and the laser

optical resolution of 78 μm per pixel, using an effective image size of 320 × 256 pixels (half frame) and a file of view of 25 × 20 mm and capturing at a frame rate of 200 Hz. The distance between camera and crankshaft surface is 200 mm with a tolerance of ±6 mm. Laser power used for unprocessed surfaces is between 5 and 10 W, and for processed surfaces with lower emissivity values, a laser power of 50 W is used at a scanning speed between 60 and 200 mm/s. For safety reasons, the workcell is made of hard metal sheets and a security switch which allows laser operation only when doors are properly closed.

10.5   Conclusions and Outlook

From the first implementations of industrial inspection robots for quality control, much progress has been made in terms of increasing the level of autonomy and making such complex systems easier to handle. The solution of complex inspection problems is already feasible today, but still requires significant effort for the implementation of a particular application. Semi-automatic methods for setting up the inspection process help to reduce this effort, but have not yet reached a level of abstraction that would allow generic solutions for whole classes of applications. The structures and examples presented in this chapter provide a basis for such generic solutions that require specific modules, such as augmented CAD data of the part, a process model for the inspection process and a 3D model of the workcell to set up a whole inspection task. Future research will enable an easy exchange of the single modules to quickly adapt an inspection system to a new task or a new inspection process and it will not take too long, before such systems are installed in industrial production lines.

References

1. Tătar O, Mândru D, Ardelean I (2007) Development of mobile minirobots for in pipe inspection tasks. MECHANIKA, 6(68):1392–1207
2. Granosik G, Borenstein J, Hansen MG (2006) Serpentine robots for industrial inspection and surveillance. In: Huat LK (ed) Industrial robotics: programming, simulation and application, pp 702, ISBN: 3 86611-286-6
3. Edinbarough I, Balderas R, Bose S (2005) A vision and robot based on-line inspection monitoring system for electronic manufacturing. Comput Ind 56:986–996
4. Woern H, Laengle T, Gauss M (2003) ARIKT: adaptive robot based visual inspection. Künstliche Intelligenz 2:33–35

5. Kuhlenkoetter B, Krewet C, Schueppstuhl T (2006) Adaptive robot based reworking system, industrial robotics: programming, simulation and applications. Low Kin Huat (ed), ISBN: 3-86611-286-6

6. Yang CC, Ciarallo FW (2001) Optimized sensor placement for active visual inspection. J Robotic Syst 18(1):1–15

7. Biegelbauer G, Vincze M, Noehmayer H, Eberst C (2004) Sensor based robotics for fully automated inspection of bores at low volume high variant parts. IEEE international conference on robotics and automation, 5:4852—4857, 26 April–1 May 2004

8. Merat FL, Radack GM (1992) Automatic inspection planning within a feature-based CAD system. Robotics Comput Integr Manuf 9(1):61–66

9. Park TH, Kim HJ, Kim N (2006) Path planning of automated optical inspection machines for PCB assembly systems. Int J Control Autom Syst 4(1):96–104

10. Scoot WR, Roth G (2003) View planning for automated three-dimensional object reconstruction and inspection. ACM Comput Surv 35(1):64–96

11. Lee KH, Park HP (2000) Automated inspection planning of free-form shape parts by laser scanning. Robot Comput Integr Manuf 16(4):201–210

12. Lu CG, Morton D, Wu MH, Myler P (1999) Genetic algorithm modelling and solution of inspection path planning on a coordinate measuring machine (CMM). Int J Adv Manuf Technol 15:409–416

13. Scott W, Roth G, Rivest JF (2002) Pose error effects on range sensing. In Proceedings of the 15th international conference on vision interface (Calgary, Alta., Canada), pp 331–338

14. Pisinger D, Ropke S (2010) Large neighborhood search. Handbook of Metaheuristics, pp 399–419

15. Dorigo M, Maniezzo V, Colorni A (1996) The ant system: optimization by a colony of cooperating agents. IEEE Trans Syst Man Cybern Part B 26(1):29–41

16. Ankerl M, Hämmerle A (2009) Applying ant colony optimisation to dynamic pickup and delivery. In: Comput Aided Syst Theory-EUROCAST, pp 721–728

17. Daniilidis K (1999) Hand-eye calibration using dual quaternions. Int J Robot Res 18(3):286–298

18. Zhao Z, Liu Y (2006) Hand-eye calibration based on screw motions. 18th International conference on pattern recognition (ICPR'06), 3:1022–1026

19. Hollerbach JM, Wampler CW (1996) The calibration index and taxonomy for robot kinematic calibration methods. Int J Robot Res 15(6):573–591

20. Pradeep V, Konolige K, Berger E (2010) Calibrating a multi-arm multi-sensor robot: a bundle adjustment approach. In: Proceedings of the International symposium on experimental robotics (ISER), Delhi India, Dec 18–21 2010

21. Strobl KH, Hirzinger G (2008) More accurate camera and hand-eye calibrations with unknown grid pattern dimensions. Proceedings-IEEE international conference on robotics and automation, pp 1398–1405

22. Eitzinger C, Heidl W, Lughofer E, Raiser S, Smith JE, Tahir MA, Sannen D, Van Brussel H (2009) Assessment of the influence of adaptive components in trainable surface inspection systems. Mach Vis Appl J, doi: 10.1007/s00138-009-0211-1

23. Smith JE, Tahir MA, Caleb-Solly P, Lughofer E, Eitzinger C, Sannen D, Nuttin M (2009) Human-machine interaction issues in quality control based on on-line image classification. IEEE Trans Syst Man Cybern 39(5):960–971

24. Eitzinger C, Thumfart S (2012) Optimizing feature calculation in adaptive machine vision systems. In: Sayed-Mouchaweh M, Lughofer E, Learning in non-stationary environments: methods and applications, Springer Science+Business Media New York, doi: 10.1007/978-1-4419-8020-5_13

25. Szeliski R (2006) Image alignment and stitching: a tutorial. Found Trends Comput Graph Vis 2(1):1–104

26. Brown M, Lowe D (2007) Automatic panoramic image stitching using invariant features. Int J Comput Vision 74(1):59–73

27. Kopf C, Heindl C, Rooker M, Bauer H, Pichler A (2013) A portable, low-cost 3D body scanning system. 4th International conference and exhibition on 3D body scanning technologies, CA USA Nov 19–20 2013
28. Schmitt R, Mersmann C, Schoenberg A (2009) Machine vision industrialising the textile-based FRP production. In: Proceedings of 6th international symposium on image and signal processing and analysis, pp 260–264
29. Palfinger W, Thumfart S, Eitzinger C (2011) Photometric stereo on carbon fibre surfaces. Proceeding of the Austrian association for pattern recognition
30. Thumfart S, Palfinger W, Stöger M, Eitzinger C (2013) Accurate fibre orientation measurement for carbon fibre surfaces. 15th International conference on computer analysis of images and patterns, pp. 75–82
31. Woodham R (1989) Photometric method for determining surface orientation from multiple images. Opt Eng 19(1):139–144
32. Johnson M, Adelson E (2011) Microgeometry capture using an elastomeric sensor. Comput Vis Pattern Recogn pp 2553–2560
33. Zhou K, Wang L, Tong Y, Desbrun M, Guo B, Shum HY (2005) Texture montage: seamless texturing of arbitrary surfaces from multiple images. Proceedings of ACM SIGGRAPH, pp 1148–1155
34. Avdelidis N, Gan T-H, Ibarra-Castanedo C, Maldaque X (2011) Infrared thermography as a non-destructive tool for martials characterisation and assessment. Proceedings—SPIE the international society for optical engineering, (8013–8039) Thermal Infrared Applications XXXIII
35. Holst G (2000) Common sense approach to thermal imaging. SPIE Volume PM-86, pp 60, ISBN: 0-8194-3722-0
36. Taib S, Jadin M, Kabir S (2012) Thermal Imaging for enhancing inspection reliability: detection and characterization, infrared thermography, Dr. Raghu V Prakash (Ed.), ISBN: 978-953-51-0242-7
37. Traxler G, Thanner P (2011) Automatisierte Wärmeflussprüfungen in der Stahlindustrie, Leitfaden zur Wärmeflussthermografie, ISBN 978-8396-0234-8
38. Ghidoni S, Minella M, Nanni L, Ferrari C, Moro M, Pagello E, Menegatti E (2013) Automatic crack detection in thermal images for metal parts. International conference on heating by electromagnetic sources (HES-13)

Chapter 11
Machine Vision Techniques for Condition Assessment of Civil Infrastructure

**Christian Koch, Zhenhua Zhu, Stephanie German Paal
and Ioannis Brilakis**

Abstract Manual visual inspection is the main form of assessing the physical and functional conditions of civil infrastructure at regular intervals to ensure the infrastructure still meets its present service requirements. In addition to this form of inspection, several novel machine vision techniques are gradually becoming available. They promise to reduce the time needed to inspect facilities and standardize the quality of the results by removing human judgment as a factor of variability. This chapter explains the origins and a representative selection of these methods and gives a sneak peak of future trends in machine vision techniques in assessing civil infrastructure conditions. This chapter starts with the introduction of the current practices of civil infrastructure condition assessments. Then, the state-of-the-art machine vision techniques available for the condition assessment of civil infrastructure are described. The benefits and limitations of each technique are discussed, and the challenges of using the techniques are highlighted. Several case studies are presented to show the effectiveness of these techniques in assessing the conditions of civil infrastructure, such as bridges, buildings, and roads.

C. Koch
Department of Civil Engineering, University of Nottingham, University Park,
Nottingham NG7 2RD, UK
e-mail: christian.koch@nottingham.ac.uk

Z. Zhu
Department of Building, Civil and Environmental Engineering, Concordia University,
1455 de Maisonneuve West EV-6.237, Montreal, QC H3G 1M8, Canada
e-mail: zhenhua.zhu@concordia.ca

S. German Paal
Applied Computing and Mechanics Laboratory, School of Architectural, Civil and
Environmental Engineering, école Polytechnique Fédérale de Lausanne, Lausanne,
Switzerland
e-mail: stephanie.german@epfl.ch

I. Brilakis (✉)
Construction IT Laboratory, Department of Engineering, University of Cambridge,
ISG-62, Trumpington Street, Cambridge CB2 1PZ, UK
e-mail: ib340@cam.ac.uk

Contents

11.1   Introduction

To ensure safety and serviceability, it is always necessary to visually inspect and assess the physical and functional conditions of civil infrastructure at regular intervals (routine inspection) or after disasters (post-disaster inspection). Typically, such inspections and assessments are performed manually by certified inspectors and/or structural experts. The inspection and assessment include the detection of the defects and damage (cracks, spalling, potholes, etc.) inflicted on civil infrastructure elements, such as buildings, bridges and road networks, and the measurements of defect properties (width, length, etc.). The visual inspection and assessment results might help the agencies decide whether the follow-up structural integrity evaluations are needed and ensure the civil infrastructure still meets its service requirements.

This chapter starts with the description of the current practices of inspecting and assessing the conditions of reinforced concrete structures and asphalt pavements. The focus is placed on answering the following questions: (1) What are the typical defects that cause damage of civil infrastructure? (2) What are the common manual procedures to detect those defects? (3) What are the limitations of manual defect detection? (4) How are the defects measured? and (5) What tools and metrics are used to assess the infrastructure condition?

Then, the state-of-the-art machine vision techniques available for the condition assessment of civil infrastructure are described. These techniques are built upon common image processing techniques, such as image thresholding, segmentation, thinning. It is shown how techniques have been tested and evaluated to inspect reinforced concrete cracks and spalling, and asphalt pavement cracks and potholes. The results indicate that machine vision techniques can be used for the damage detection and measurement of reinforced concrete structures and the distress detection and measurement of asphalt pavements.

11.2   Current Practices

11.2.1   Routine Inspections

11.2.1.1   Bridge Inspections

Defect detection

Highway bridge inspections are carried out manually by certified inspectors following the established standards and manuals (i.e., National Bridge Inspection Standards and AASHTO Manual on Bridge Evaluation). Before going onto a bridge, the inspectors prepare sketches and note templates for references throughout the inspection [53]. During inspection, they record the actual bridge conditions by observing existing defects, such as cracks, that lay on primary bridge components, such as decks, exterior beams, and piers (Fig. 11.1).

Such manual inspections have several inherent limitations. First, they are inefficient. Reviewing previous reports, preparing inspection plans, and collecting and analyzing field data require a large amount of inspector hours [10, 30], which has proven costly over time. In Tennessee, in 2006 and 2007, the annual cost of bridge inspection was $7.6 million; considering the Tennessee Department of Transportation lost $22 million in federal bridge funds in the last two years, some bridge inspections and maintenance projects have to be pushed back [61]. In addition, the results from manual inspections are subjective and highly variable. An on-site survey conducted by FHWA Non-Destructive Evaluation Center (NDEVC) indicated that manual inspections were completed with significant variability [45, 49]. Finally, the current guidelines need considerable updating with regard to health monitoring and nondestructive testing techniques, as well as in regard to the proper integration of that data into the rating process [21].
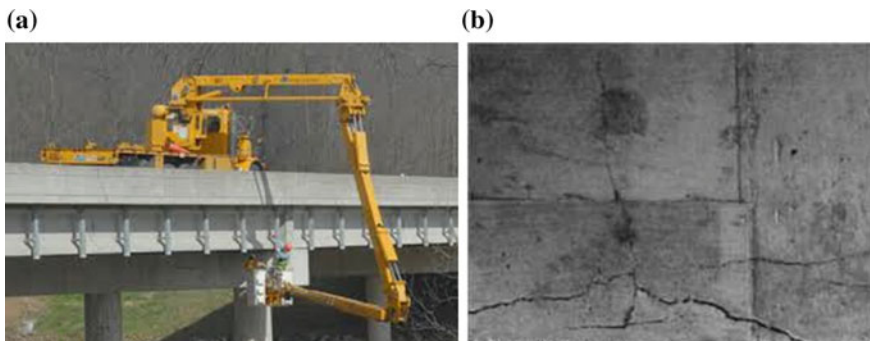
Fig. 11.1   a Routine bridge inspection [43] and **b** typical bridge surface defects [1]

Defect measurements

As mentioned before, cracks are one of the common defects inflicted on concrete structural elements, such as columns, beams, walls. Therefore, it is always necessary to inspect the cracks at routine inspections. During the crack inspection process, the first important step is to identify the location and severity of cracking. Then, the cause and effect of the cracking are established. Whether or not the observed cracks are indicative of potential structural problems could be determined, considering the loading conditions of the structural elements. This way, the strategy for the repair of the cracks could be made in order to maintain the structural integrity and/or the exterior structural appearance.

Typically, the cracking locations and severity on the concrete structural elements are identified through direct manual observations. However, the observation results are qualitative not quantitative. This limitation could be partially overcome by placing a grid on the element surfaces to accurately locate the cracks. Also, the crack widths can be measured using a crack comparator, which is a small, handheld microscope with a built-in measuring scale on the lens [3] (Fig. 11.2). The movement of a crack could even be monitored with a crack monitor, which could give a direct reading of crack displacement and indicate the maximum movement during the monitoring period [3] (Fig. 11.3). If possible, the crack depth can be measured by inserting a fine wire or gage.

When properties (location, width, depth) of a crack are retrieved, the type of the crack could be defined accordingly. For example, crack can be classified as longitudinal, transverse, or diagonal, based on its specific direction. If the crack width is considered as the classification criteria, the crack can be classified into three ranges: fine (generally less than 1 mm), medium (between 1 and 2 mm), and wide (over 2 mm) [3].

Severe crack conditions need to be visually documented, so that the current and previous cracking conditions can be compared. Typically, the photographs of the

Fig. 11.2 Comparator for measuring crack width [3]
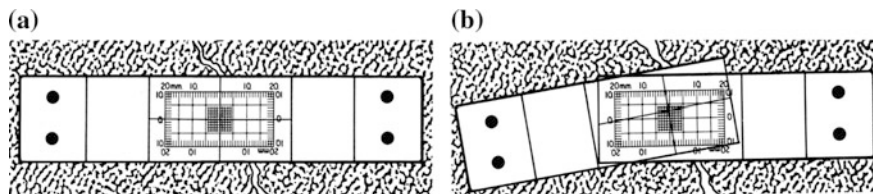
Fig. 11.3 Crack monitor: **a** newly mounted monitor, **b** monitor after crack movement [3]

cracks are taken. The photographs do not only recording crack conditions, but also include the carbonation of the surfaces around the cracks, the spalling along the crack edges.

All the activities and measurements mentioned above are performed in a routine visual inspection, which is one of the first steps in any comprehensive assessment of a concrete structure [48]. The visual inspection results may indicate the necessity of performing other further assessments. For example, during assessment of existing nuclear safety-related concrete structures, it is pointed out that further assessments are not necessary, when concrete surfaces exposed for inspection meet such surface conditions: (1) popouts and voids are less than 20 mm in diameter or equivalent surface area, (2) spalling is less than 100 mm in any dimension, and (3) cracks less than 0.4 mm in the maximum width [4].

Although visual inspection is a powerful tool and widely used in the construction industry, it has been identified with several limitations. At first, its effectiveness heavily depends on the knowledge and experience of the inspector in structural engineering, concrete materials, and construction methods. This experience-based inspection is not always reliable, since inspectors may interpret the results of the finished product differently. In addition to the subjective nature, the manual inspection process is time-consuming, especially if the concrete structure under inspection is complex [9]. The requirement of experienced inspectors also poses a big challenge for the construction industry, which is facing the pressing shortage of experienced and highly trained inspection personnel.

Condition assessment

The manually detected defects, such as cracks (e.g., flexural, shear, vertical and bond cracks), loss of cover and spalling, and corrosion and efflorescence, are used to rate bridge conditions. So far, there are two rating systems that can be adopted. The FHWA Recording and Coding Guide [19] defines one system which uses the National Bridge Inventory (NBI) zero to nine scale for rating a bridge deck superstructure and substructure. The second system, the PONTIS rating system, uses a set of three to five condition states to describe the condition of approximately 160 bridge elements, such as columns, girders, slabs, and trusses [54]. Both rating systems are built on qualitative definitions. For example, in the NBI rating system, a bridge is rated at "Fair Condition" (5) when all of its primary components are sound but may have minor cracking, spalling, or scour [19].

11.2.1.2 Asphalt Pavement Inspections

Distress detection

The roadway network of a developed country contains thousands kilometers of pavement. A huge part of this network consists of bituminous (asphalt) pavements ranging in condition, age, and performance. As reported by ASCE, poor road conditions lead to excessive wear and tear on vehicles as well as an increased number of delays and crashes [8]. Several road network maintenance programs have been established in order to monitor the ongoing performance of the road network, to predict future pavement conditions, to assess long-term needs, to support investment planning and decision making, and to identify rehabilitation and maintenance treatment (e.g., [22]). Within these programs, pavement surface condition assessment is identified as a key component that requires reliable and good-quality measurements on pavement distresses such as cracks, potholes, and patches. The process of pavement surface condition assessment is usually divided into parts: data collection, distress detection, distress measurements, and distress assessment. In current practice, the first is to a large extent automated, while the latter three are mostly performed manually.

Sophisticated digital inspection vehicles collect pavement data using several sensors, in particular downward-looking video cameras for surface imaging, optical sensors for distance measurements, laser scanners for longitudinal and transverse profiling, additional ultrasonic sensors for rutting detection, and accelerometers for roughness measurements [23, 43]. Sensing the pavement surface with such an inspection vehicle at speeds up to 60 mph (100 km/h) is currently the most effective approach, which is quickly replacing traditional manual low-speed and manual methods [51]. A typical vehicle is depicted in Fig. 11.4a.

Due to the high costs, the number of inspection vehicles within a certain area is very limited. This results in a survey cycle of one year to four years duration for critical roadways, such as highways, and many years or complete negligence for all
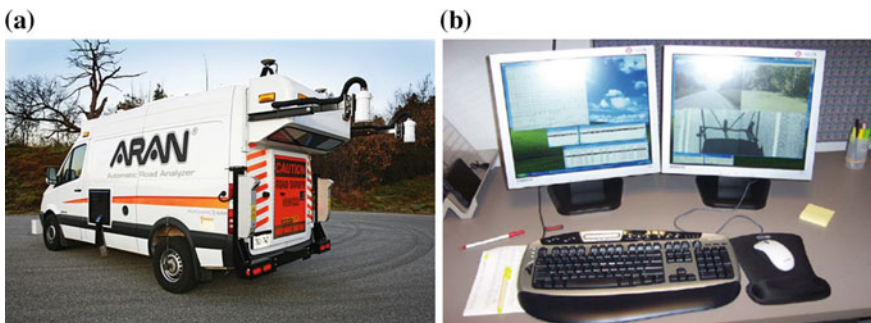
Fig. 11.4 Pavement condition assessment: **a** Automatic road analyzer (ARAN) at Fugro Roadware Inc. [23] for data collection. **b** Workstation for manual or semiautomated defect detection and assessment [43]

other local and regional roads. Furthermore, only a 10 % sample (first 500 ft. of each mile) of the pavement miles is actually inspected and used to perform the assessment rating [43]. The time associated with sensing the whole highway network of a federal state does not allow for repeated measurements to improve the accuracy and reliability of a survey. In cases where there are no digital inspection vehicles available, pavement condition surveys are still manually conducted either by foot or from a van at a speed of 8–15 mph (10–25 km/h) using data collection assistance software [51].

Basically, there are five major surface distress categories, which are illustrated in Fig. 11.5:

- Cracking [transverse, longitudinal (Fig. 11.5a), block/alligator cracking (Fig. 11.5b)];
- Patching (Fig. 11.5c);
- Potholes (Fig. 11.5d);
- Rutting and shoving (Fig. 11.5e); and
- Bleeding and raveling (Fig. 11.5f).

The digital pavement surface image and video data captured by the downward-looking cameras are predominantly manually reviewed and analyzed by technicians using computer workstations (Fig. 11.4b). Sitting in front of two or more screens, they visually detect distresses based on their own experience and distress manuals (e.g., [20, 42]). Visual distress identification manuals define categories of type, severity, and extent of distress. While some types of distress can be easily observed visually, such as cracking, patching, potholes and raveling, rutting can only be identified based on the additional 3D information, such as profile data
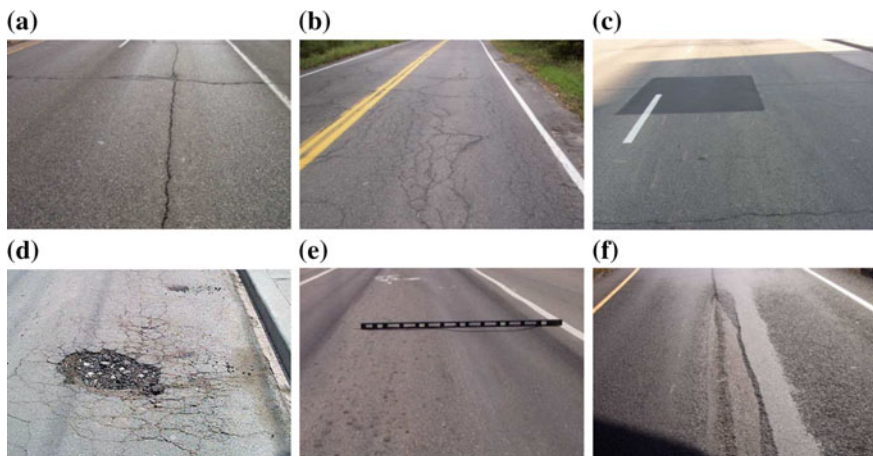
Fig. 11.5 Examples of asphalt pavement surface distress [42]. **a**, **b** cracking, **c** patching, **d** potholes, **e** rutting, **f** raveling

captured by laser scanners or ultrasonic sensors. Besides solely manual detection, there are a few software packages available that semiautomatically assist in the detection and classification of surface distress, such as cracking (e.g., [23]).

Distress measurements

Visual pavement distress identification manuals do not only define categories of type, they also define units of measure. However, these units vary between different countries, and federal states. Table 11.1 presents examples of defined units related to each type of pavement surface distress.

Within manual condition surveys, inspectors measure the extent of distress manually (manual counting, using a ruler). In case of distress detection in available pavement surface images and videos, the data can be registered according to scale, such that measures such as number per length, length, lineal feet, area, and area ratio can be determined automatically. Integrating all detection and measurement results usually leads to a so-called distress map that graphically depicts the type and the extent of pavement distress for a certain road segment (Fig. 11.6).

Table 11.1 Units of asphalt pavement distress measures

Type of distress	US DOT (FHWA) [20]	Minnesota DOT [42]	German DOT [24]	UK DFT [57]
Cracking	Number, length, area	Number, lineal feet	Area ratio	Number, length, area
Patching	Number, area	Lineal feet	Area ratio	Depth, area
Potholes	Number, area	Number	(area ratio)	Number
Rutting, shoving	Depth, number, area	Lineal feet,	Depth	Depth
Bleeding, raveling	Area	Lineal feet	Area ratio	Area
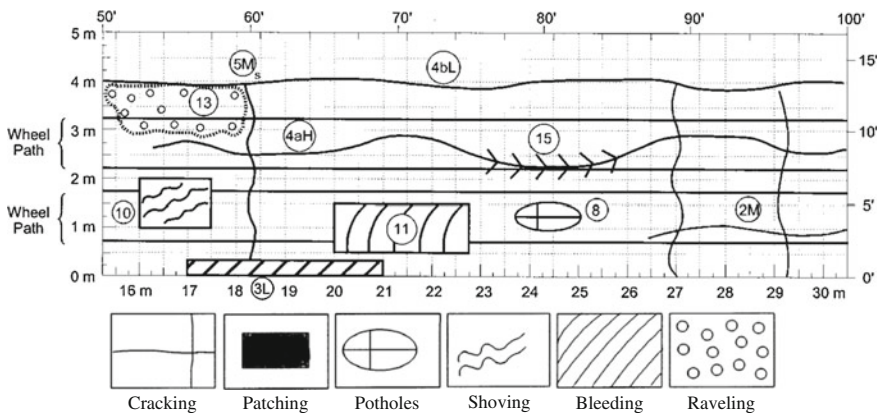
Fig. 11.6 Example of a distress map for a 30 m pavement segment according to FHWA [20]

Condition assessment

In general, based on the extent of identified distresses and the equivalent severities, a surface distress index is calculated for a certain pavement segment, e.g., 100 m, or 500 ft. For this purpose, visual distress identification manuals provide details on how to map the distress measures (cp. Sect. 11.2.2) to severity levels, such as low, medium, and high (e.g., [24, 42]). The amount of each distress (e.g., number of cracks) is converted into a distress percentage. Additionally, individual types of distress are weighted differently when calculating the total distress index. For example, a certain distress type at a certain severity level yields a certain weighting factor (e.g., medium transverse cracking: 0.1). Multiplying the distress percentage with the weighting factor results in the individual weighting distress that is summed up to yield the total weighted distress. Finally, the total weighted distress can be directly mapped onto a surface distress index, which is equivalent to the surface rating.

In order to determine an overall Pavement Quality (Condition) Index the surface distress index (surface rating) is combined with a roughness index, which is derived from transverse and longitudinal profiles derived from 3D laser scans. Finally, the Pavement Quality Index can be transformed into a value for the remaining service life in years [43]. Moreover, the detailed assessment results support the decision making regarding an adequate repair strategy.

11.2.2 Post-disaster Inspections

11.2.2.1 Damage Detection

The safety evaluation of buildings in the event of an earthquake, for example, is twofold: Buildings must be evaluated prior to entry of emergency search and rescue teams, and again, the same buildings must be evaluated prior to reentry by occupants. Each of these safety evaluations is currently performed by certified inspectors or structural engineers. These specialists follow existing procedures established by local (national) government authorities in most cases; however, worldwide, several countries adhere to those guidelines established by US-based entities such as the Federal Emergency Management Agency (FEMA) (emergency response) and the Applied Technology Council (ATC) (occupancy) [7, 18]. The specialists must make an assessment based on their experience and knowledge coupled with their visual observation of the damage inflicted on the load-bearing members of a structure (Fig. 11.7).

In the case of evaluating the safety of post-earthquake structures for emergency responders, structural specialists are considered the most prepared personnel to deal with all aspects of the built environment in urban areas [5]. The involved structural specialists are responsible for identifying potential structural hazards and

COLUMN FAILURE SO THAT
FLOOR & ROOF ABOVE ARE
NOW DRAPED BETWEEN ADJACENT
COLUMNS AND PULLING ON REST
OF STRUCTURE

LOOSE H.V.A.C. EQUIP
AND/OR WATER TANK

LOOSE SIGNS OR WALL
PANELS & ORNAMENTS
MAY FALL

BROKEN
ELECTRIC
LINES

CRACKED FL
AT COLUMN
(PUNCH SHEAR)

CONCRETE
FLOOR OR
WALL PIECE
HANGING BY
REBAR

ACME
HOSPITAL

BROKEN GAS
& WATER LINE

BADLY CRACKED
CONCRETE WALLS

CONCRETE BROKEN
OFF REBAR CAGE
AT JOINT BETWEEN
COLUMN & FLOOR

CONCRETE MISSING FROM
INSIDE REBAR CAGE
(EMPTY BASKET)

BADLY CRACKED INFILL
WALLS OF UNREINFORCED
MASONRY MAY FALL OUT

Fig. 11.7 Basic collapse patterns and check points for a concrete frame building as noted in the structural collapse technician course student manual [18]

monitoring the structure for condition changes during rescue and recovery operations [18]. Search and rescue teams can only enter the buildings that are determined structurally stable by structural specialists. However, several collaboration-related problems between civil engineers and other organizations involved in disaster relief efforts have been recognized, including the lack of coordination, information sharing, trust, and communication [36].

In the case of evaluating the safety of post-earthquake structures for occupants, the ATC-20 code outlines three procedural levels: rapid evaluation, detailed evaluation, and engineering evaluation [7]. Rapid evaluation is typically based on a very rough inspection of the exterior structure; detailed evaluation is a more thorough visual inspection of the structures interior and exterior; and in an engineering evaluation, engineers supplement the procedures carried out in the rapid and detailed evaluations by investigating the safety of a damaged structure from construction drawings and new structural calculations [46] The purpose of the rapid evaluation is to quickly identify apparently "unsafe" or "safe" buildings after an
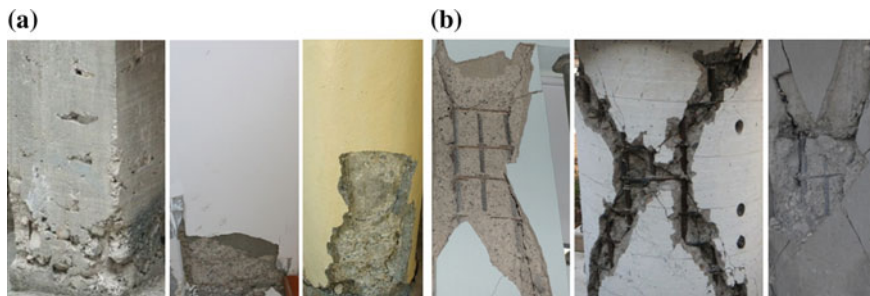
Fig. 11.8 Examples of spalled concrete images: **a** spalling that only exposes the cover concrete; and **b** spalling that exposes reinforcement

earthquake. A building is regarded as unsafe if it partially collapses or its stories lean severely [7]. For example, a cast-in-place concrete building is regarded as unsafe when any of three conditions exists in the columns: (1) buckled or fractured reinforcement; (2) massive concrete spalling and exposure of longitudinal reinforcement (Fig. 11.8); or (3) wide diagonal cracks through the column depth [7].

11.2.2.2   Damage Measurements

Assessments can take months, due to the large number of buildings that earthquakes affect and the comparatively small number of available inspectors. Reducing this time is possible by automating the laborious steps of the assessment process. One such step is the measurement of damage. Specifically, certain useful damage properties need to be retrieved for structural integrity assessment. "Useful damage properties" can be defined as those properties which can be easily translated to an indicator of the overall damage of the building component and/or structure as a whole. Figure 11.9 depicts a typical evaluation form that is created within the manual procedure of RC column inspection for post-earthquake safety assessment.

11.2.2.3   Condition Assessment

The buildings that cannot be determined as "safe" or "unsafe" within the rapid evaluation procedure are further assessed in the detailed and engineering evaluations. Here, the severity and extent of damage to the structural and non-structural elements throughout the building is observed, measured, and recorded. The purpose of all of these procedures is to ultimately designate the buildings as falling into one of three categories: (1) RED-TAG: imminent threat to lifesafety; (2) YELLOW-TAG: risk from damage but not imminent threat to lifesafety; and (3) GREEN-TAG: safe for entry and occupancy as earthquake damage has not significantly affected the safety of
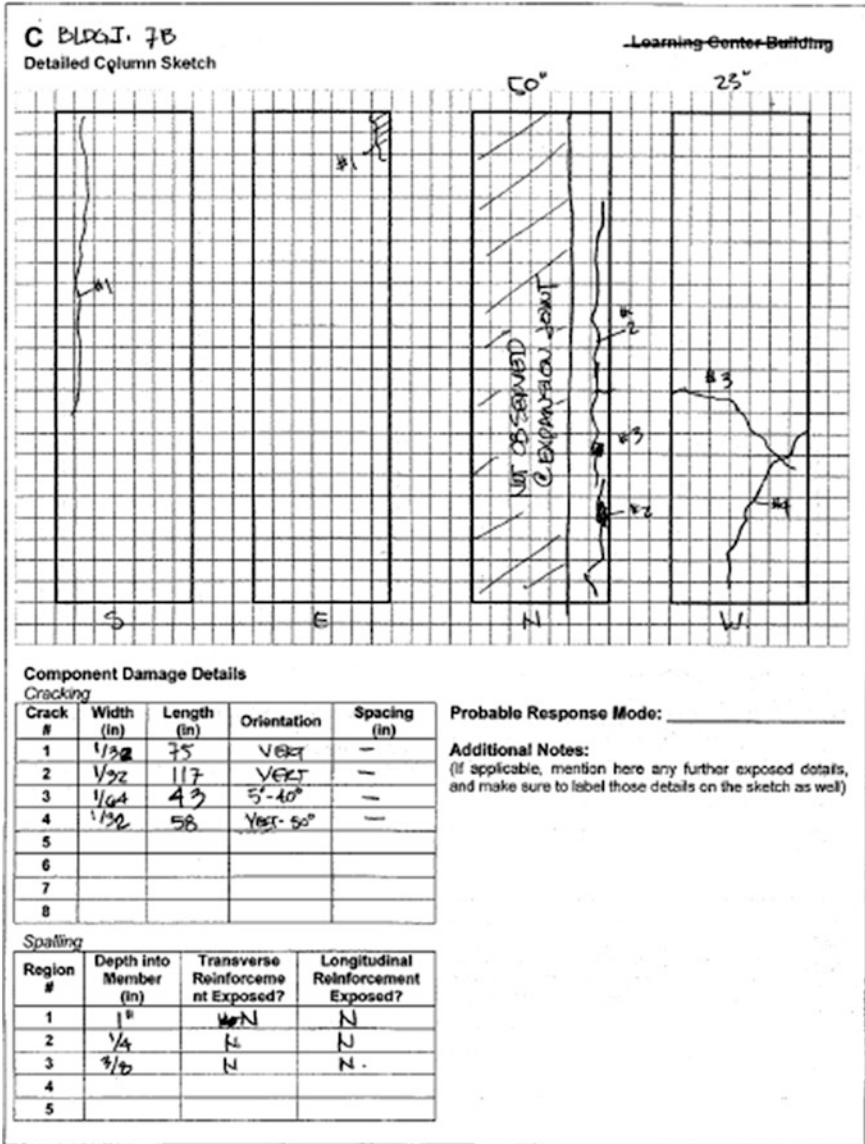
Fig. 11.9 Manual evaluation form for columns in post-earthquake safety assessment: cracking and spalling

the building. These categorical guidelines are inherently subjective and require significant judgment on the part of the inspector. Table 11.2 presents exemplified building assessment criteria for RC columns.

Table 11.2 Draft SEAONC disaster emergency service committee building assessment criteria [7] with focus on columns

Red (5, 6)	Yellow (3, 4)	Green (1, 2)
• Significantly out of plumb	• No apparent instability hazard	
• Moderate buckling	• Minor cracking	
• Severe cracking		
• Signs of yielding		
• Falling hazard		
• Damaged, so as to no longer be providing support to level above		

11.3 Machine Vision Techniques

11.3.1 Defect Detection for Reinforced Concrete (RC) Structures

Significant progress has been made in the general field of computer vision, which can be applied to damage detection for reinforced concrete structures. The majority of the progress in automated damage detection in concrete structures has focused on cracking. However, concrete palling has been accepted as an important indicator of significant damage to structural elements during post-disaster events such as earthquakes.

11.3.1.1 RC Cracks

In order to detect the concrete cracks from images or videos, many machine vision-based methods have been created. Typically, the methods make use of existing vision techniques, such as principal component analysis (PCA), edge detection, Fourier transform, and wavelet analysis. This way, the cracks can be detected with little human interventions.

Existing machine vision-based methods can be classified into two categories based on the results that could be achieved. The methods in the first category can only recognize whether or not the image or video of a structural element surface contains a crack (i.e., crack presence recognition). Abdel-Qader et al. [2] proposed a PCA-based method, where the presence of the cracks in a bridge surface image can be identified. Under the method, an image is first divided into sixteen square blocks. Each block is filtered by linear feature detectors and then projected onto dominant eigenvectors which are pregenerated. The projection result is compared with the projection results of the training data. This way, whether or not the block contains cracks can be identified. Instead of using PCA-based method, Liu et al. [41] created

a crack classification system. Under their system, a support vector machine (SVM) is used to classify each region in an image as: "crack," "non-crack," or "intermediate."

The methods in the second category can not only identify the presence of cracks in an image, but also locate crack points in the image and produce a crack map. The crack map is a binary image, where the crack points are typically marked white and non-crack points (surface background) are black. A lot of image processing techniques can be used to produce crack maps, including wavelet transforms, thresholding, and edge detection. For example, Cheng et al. [17] detected cracks in an image by simply thresholding the concrete surface image, where the threshold value was determined based on the image's mean and standard deviation. Abdel-Qader et al. [1] compared the effectiveness of four edge detection techniques (the Canny edge detector, Sobel edge detector, Fourier transform, and fast Haar transform). According to their test results, it was found that the fast Haar transform was most reliable. Sinha and Fieguth [50] introduced two local crack detectors, which considered relative statistical properties of adjacent image regions. Similarly, Iyer and Sinha [29] designed morphology-based crack filters, which were constructed with linear structuring elements. Also, Yamaguchi and Hashimoto [62] proposed a type of scalable percolation-based method that considered crack connectivity.

The crack map only contains the locations of the crack points in an image. It is unknown how these crack points are connected to form different cracks. Also, it is unclear about how to get the specific crack properties, such as crack length, orientation, and maximum width. In order to address these problems, Yu et al. [63] relied on a graph search to calculate the length, thickness, and orientation of concrete cracks, when the start and end points of the cracks are manually provided. Chae et al. [12] developed an artificial neural network, which could retrieve crack properties.

The recent work related to the retrieval of crack properties from the crack map could be found in Zhu et al. [65]. In their work, the crack map was first thinned with a binary image thinning algorithm [44] to produce crack skeleton points. Also, a Euclidean distance transform [11] was used to calculate the distance of each crack skeleton point in the map to the crack boundaries. Moreover, the crack topological configuration was identified by checking the connectivity of the crack skeleton points in the map. Based on the crack skeleton points, distance values, and topological configurations, it was possible to retrieve the properties of a crack, including length, orientation, and width information (Fig. 11.10).

11.3.1.2   RC Spalling

Although no work prior to that presented by German et al. [25] has been performed in the area of spalling detection and property retrieval, prior research efforts in rust detection and property retrieval are relevant due to the similarities in the two damage types (heavily chaotic and distinct in color). A myriad of methods use grayscale images to detect the existence of rust defects NFRA [14], BE-ANFIS
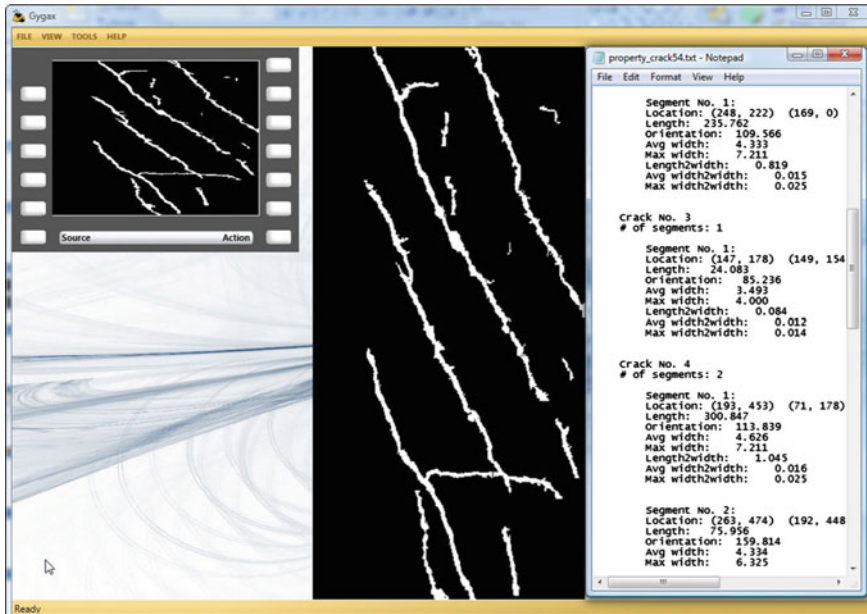
Fig. 11.10 Example of a crack with the properties retrieved [65]

[15]. In order to work in the gray scale, the color image is converted for processing, causing loss of information which then results in miscalculation of uneven background spots as rust [16]. Lee et al. [38] developed a method in rust detection which determines whether or not rust defects exist in a given image based on the statistical data acquisition and multivariate analysis processing of the image in color. Lee [39] also proposed a method based on the comparison between eigenvalues of defective and non-defective images for the purpose of detecting the presence of rust in an image. Dealing with the color space rather than the grayscale space for rust/corrosion images eases error pertaining to non-uniform illumination as well as preserves the original information of the image, allowing for more accurate results. Chen et al. [16] proposed a combination of Fourier transform and SVM-based method (SVMRA) which is capable of handling non-uniform illumination by using the L*a*b* color space. Their method can effectively recognize rust in various paint colors and uneven illumination schemes, and the percentage of rust pixels is calculated in a classification stage. However, the processing time of SVMRA is not as fast as some methods in grayscale detection.

German et al. [25] created a novel method of detecting regions of spalling and retrieving key properties pertaining to those spalled regions on RC columns (Fig. 11.11a). Considering the visual appearance of spalled concrete regions, four main distinctive visual characteristics have been identified: (1) a highly textured region, (2) ribbed texture along the surface of any potentially exposed reinforcement, (3) the distinct color of steel reinforcement, and (4) the width of the

(a)   **(b)**
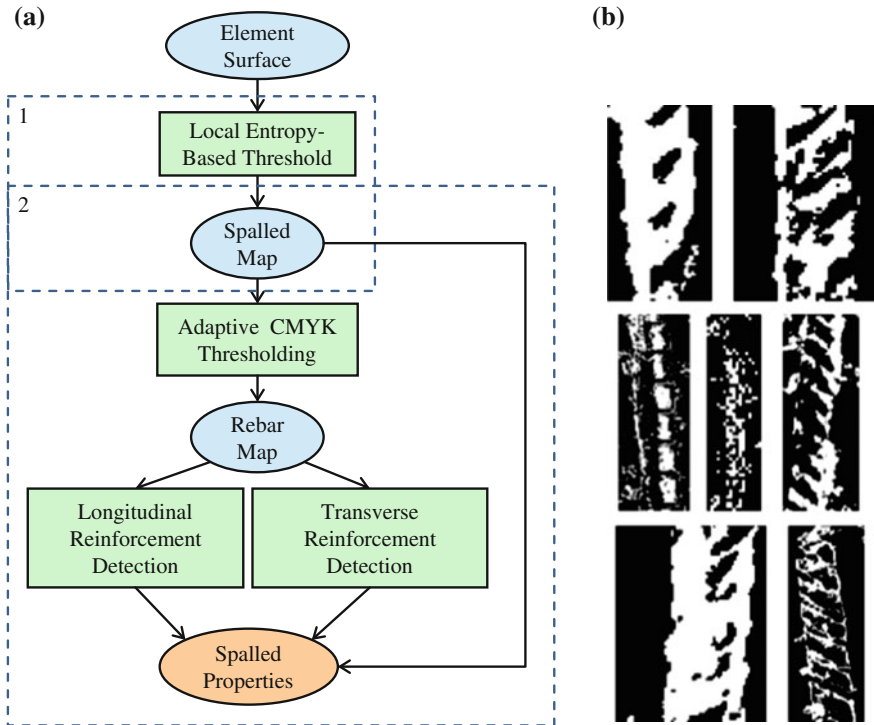
Fig. 11.11 Spalling detection and property retrieval method overview [25]. **a** *1* Detection, *2* property retrieval. **b** Vertical templates

longitudinal/transverse reinforcing steel should be significantly less than the width/height of the column. Based on these observations, image segmentation, template matching (Fig. 11.11b), and morphological filtering have been identified as the tools needed to achieve detection and property retrieval (Fig. 11.12). With respect to spalling on reinforced concrete (RC) columns, the key properties considered are as follows: (1) the length of the spalled region along the column, represented by two dimensions—LS, the total distance of the spalled region along the longitudinal direction of the structural element; and LT, the distance between extreme exposed transverse reinforcement bars, and (2) the extent (or depth) of the spalling into the column, classified as one of the following: (a) no spalling, (b) spalling of cover concrete only, (c) spalling exposing transverse reinforcement, (d) spalling exposing longitudinal reinforcement, or (e) crushing of the core concrete/spalling into the core of the element. These are commonly used as characteristics in consideration of the extent of damage to RC structural elements for routine and rapid inspection.

First, the region of interest (ROI) is determined by way of a local entropy-based thresholding algorithm (Fig. 11.12b). The ROI consists of the total spalled region,
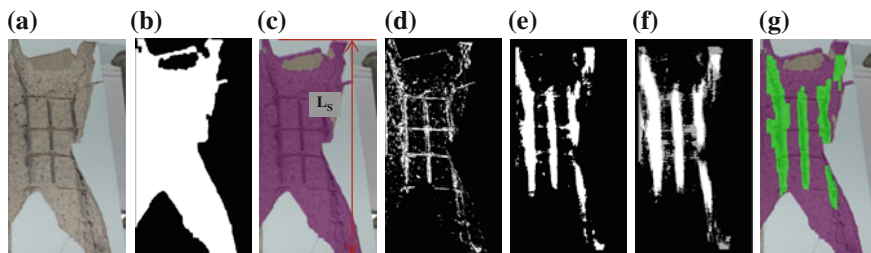
Fig. 11.12 Spalling detection and property retrieval results: **a** original image; **b** local entropy-based threshold; **c** morphological operations and length retrieval; **d** adaptive threshold on single CMYK channel; **e** sample template matching; **f** combination template matching; **g** overlayed rebar map on spalled map. [25]

including areas of cracking and/or exposed reinforcement. At this point, a combination of morphological (opening/closing/hole-filling) operations is applied to fill in the ROI and retrieve the "spalled map" (Fig. 11.12c). Then, a global adaptive thresholding algorithm is applied throughout the ROI in each of the four channels of the CMYK color space of the image to differentiate the background (damaged or undamaged concrete pixels) from the foreground (relative regions of exposed reinforcement) (Fig. 11.12d). Following this step, exposed longitudinal reinforcing steel is identified by way of a template matching algorithm (Figs. 11.11b and 11.12e–f), and the result is de-noised with further aid from morphological opening and closing operations (Fig. 11.12g). The method was implemented into the prototype tested using a database of 70 damaged RC column images. The average precision and recall for the method in spalling detection were 81.1 and 80.2 %, respectively. With the existing method, the error in the measurement of the length of the spalled region along the columns longitudinal axis (LS) was 4.22 %. With respect to the spalled region depth retrieval, at this time, the method measures this property by classifying the element into one of the following three categories: (1) no spalling, (2) cover spalling only, or (3) spalling exposing longitudinal reinforcement. The recall for this classification and property retrieval procedure is 81.5 %.

11.3.2   Distress Detection for Asphalt Pavements

In the past two decades, a huge amount of research was done in automating pavement distress detection, assessment, and repair. The most popular approaches are based computer vision algorithms that operate on 2D images to recognize, classify, and measure visible pavement surface defects, such as cracks and potholes.

11.3.2.1 Asphalt Cracks

As pavement cracking is the most common distress type, several efforts have been made with specific regard to crack detection and assessment, in particular real-time crack analysis [28, 59], crack classification [52], crack depth estimation from vision Amarasiri et al. [6], and even automating crack sealing [26, 33].

Image segmentation methods are commonly used for asphalt crack detection. Tsai et al. [55] have critically assessed available distress segmentation methods for crack detection and classification, namely statistical thresholding, Canny edge detection, multiscale wavelets, crack seed verification, iterative clipping methods, and dynamic optimization-based methods. Thresholding methods are iterative segmentation procedures that classify pixels into two classes (crack or background) based on statistical measures and probability updates [37]. As the famous Canny edge detector is the best edge detector among traditional edge detection methods, it is widely applied for pavement crack detection. However, its main disadvantage is the low accuracy in severity-level detection as it usually overestimates the distress boundaries [55]. Multiscale wavelets use various spatial and spectral image resolutions to analyze and detect singularities, such as crack edges. However, finding the adequate scale to segment the image properly is challenging [55]. A different method is crack seed verification, for example, as proposed by Huang and Xu [28]. Under this method, the image is divided into cells that are classified as crack or non-crack cells based on the contrast of each element to its neighbor. Although this method works in real time, it is not robust enough for the average image quality in practice as it is hard to find a universal contrast threshold [55]. Iterative clipping methods subdivide the image into tiles, for each of which a mean value is defined as the clipping value. Iteratively, the clipping value is updated and the difference to the mean value defines the abort criterion. This method is efficient in good lighting condition, but its performance drops in case of shadows [47]. Dynamic optimization-based segmentation methods make use of heuristic optimization procedures in order to calculate a threshold to characterize global indication characteristics (e.g., the crack shape). This procedure yields very good results for pavement crack detection, but suffers from quite long processing times [55]. Some crack detection results of the methods described above are depicted in Fig. 11.13.

Although many research studies have addressed this problem and a few commercial software packages exist [23], fully automating pavement distress detection and classification in a real-time environment has remained a challenge [55, 60]. They have concluded that the methods under consideration performed differently well according to varying lighting conditions, shadows, and crack positions and that there is no comprehensive, robust, and real-time segmentation method available.

The crack detection results, as illustrated in Fig. 11.13, reveal the location of crack pixels in an image, the so-called crack map. In order to identify crack objects and to classify them as transverse, longitudinal, or alligator crack, connectivity analysis has to be performed. This is ongoing research, however, first results can be found in, for example [56].
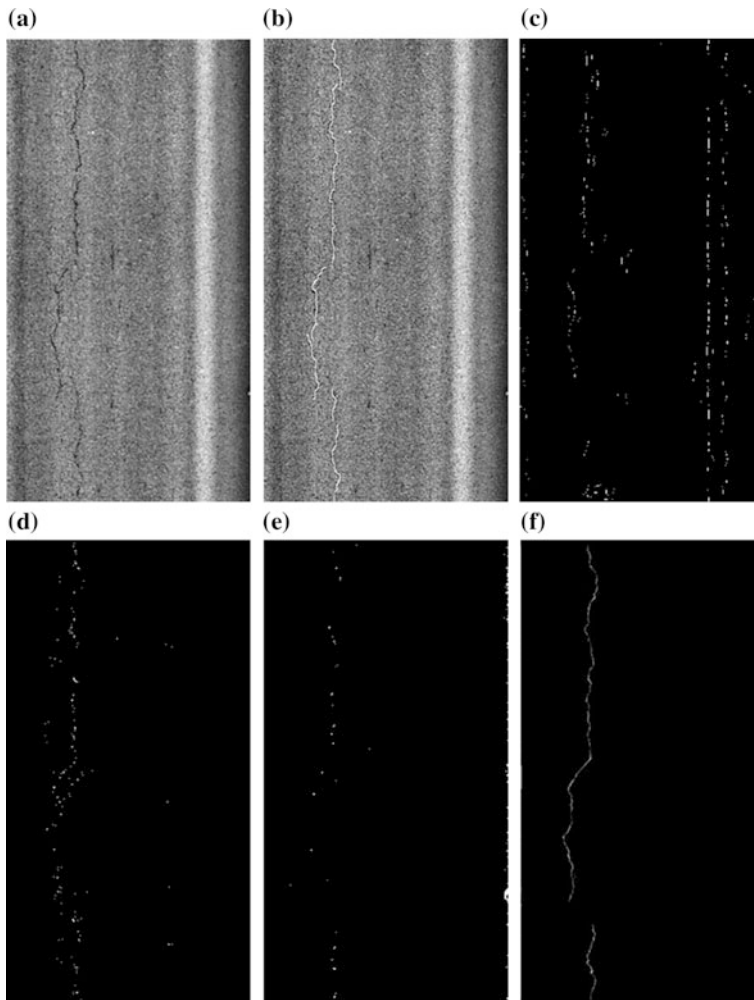
Fig. 11.13 Examples of pavement crack detection results of different methods according to Tsai et al. [55]; **a** original image, **b** ground truth, **c** Canny edge detection, **d** crack seed verification, **e** iterative clipping, **f** dynamic optimization

11.3.2.2 Asphalt Potholes

Current research efforts in automating the detection of potholes can be divided into 3D reconstruction-based, vibration-based, and vision-based methods. Detection methods that are based on a 3D reconstruction of the pavement surface rely on 3D point clouds provided by laser scanners or by stereo-vision algorithms using a pair of video cameras. Also there are hybrid systems available that use digital cameras to capture consecutive images of lines projected by infrared lasers [40].

A stereo-vision-based surface model for comprehensive pavement conditioning has been proposed by Wang [58] and Hou et al. [27]. With the availability of a 3D point cloud, Chang et al. [13] have presented a clustering approach that can quantitate the severity and coverage of potholes and Jiaqiu et al. [31] have created a method for identifying, locating, classifying, and measuring sag deformations such as potholes and depression. The drawbacks of stereo-vision-based approaches are that they require a complete 3D reconstruction of the pavement surface and that the procedure of matching points between the two views is quite challenging due to the very irregular texture and color of the pavement surface.

A different approach is based on 2D vision (image or video data) since potholes usually are significantly different from the pavement background surface. Karuppuswamy et al. [32] have presented a solution to the detection and avoidance of simulated potholes integrating a vision and motion system. Although this approach can detect potholes in the center of a lane, it relies on computer-generated (simulated) potholes that are larger than 2 ft. in diameter and white in color. However, these are very simplified assumptions that do not reflect realistic pavement conditions.
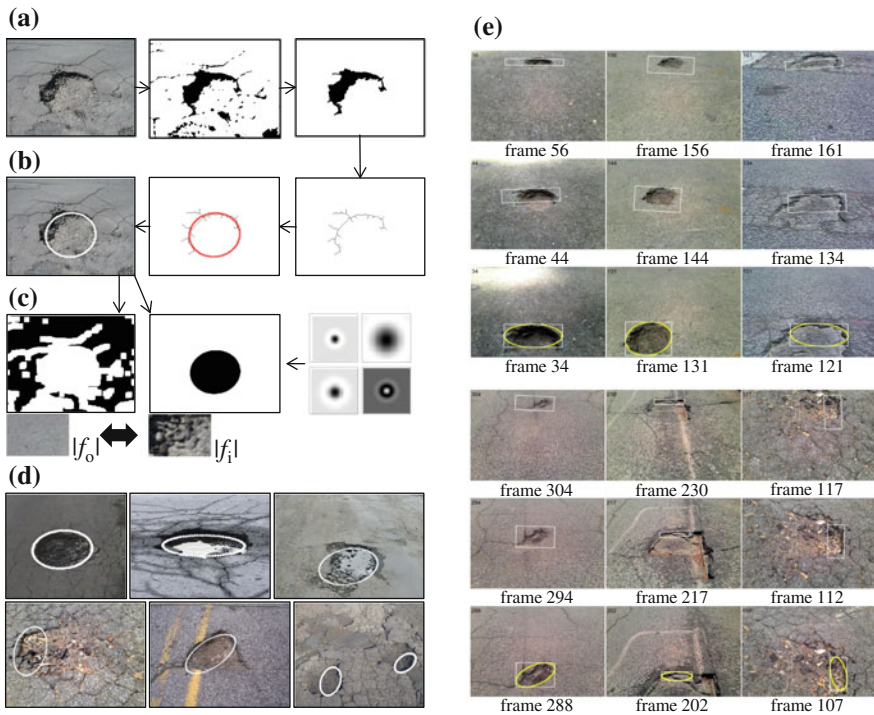
Fig. 11.14 Pothole detection method and experimental results. **a** Image segmentation. **b** Shape extraction. **c** Texture extraction and comparison. **d** Detection results in images. **e** Detection and tracking results in videos

Another vision-based pothole recognition approach has been presented in Koch and Brilakis [34]. In this method, a pothole is detected based on three visually distinctive characteristics: surrounding shadows, elliptic shape, and grain surface texture. According to this, the method subsequently performs image segmentation, shape approximation, and texture comparison. In the image segmentation step, histogram shape-based thresholding is applied to divide the image into defect and non-defect pavement regions using the triangle algorithm proposed by Zack et al. [64]. Following that, morphological thinning and elliptic regression are used to approximate the pothole shape. Finally, spot filter responses are used to compare the surface texture of the pothole candidate region and the non-defect pavement region. In case the region inside the pothole candidate is coarser and grainer than the one outside, the region is as identified as a pothole. This method has been extended with regard to pavement video processing [35]. The enhanced method recognizes and counts potholes in a sequence of pavement images using the described pothole detection method. In addition, the representative non-distress texture template is incrementally updated and a vision tracking algorithm is used to reduce computational effort and improve detection reliability. Figure 11.14 illustrates the method and its results.

11.4 Summary

It is always necessary to visually inspect and assess the physical conditions of civil infrastructure at regular intervals (routine inspection) or after disasters (post-disaster inspection). Typically, such inspections and assessments are performed manually by certified inspectors and/or structural experts. The inspection and assessment include the detection of the defects and damage (cracks, spalling, potholes, etc.) inflicted on the civil infrastructure and the measurements of their properties (width, length, etc.). The visual inspection and assessment results might help the agencies decide whether the follow-up structural integrity evaluations are needed and ensure the civil infrastructure still meets its service requirements.

The manual visual inspections and assessment methods have been identified with several limitations, including the subjective nature of inspection and assessment results and the time-consuming and labor-intensive inspection and assessment process. In order to overcome these limitations, the studies of using machine vision techniques to facilitate the visual inspection and assessment have been initiated and developed. This chapter introduced the state-of-the-art machine vision techniques for the inspection and assessment of reinforced concrete structures and asphalt pavement.

Specifically, this chapter has started with the description of the current practices of inspecting and assessing the conditions of reinforced concrete structures and asphalt pavements. The focus has been placed on answering the following questions: (1) What are the typical defects that cause damage of civil infrastructure? (2) What are the common manual procedures to detect those defects? (3) What are the limitations of manual defect detection? (4) How are the defects measured? and (5) What tools and metrics are used to assess the infrastructure condition?

Then, the state-of-the-art machine vision techniques available for the condition assessment of civil infrastructure have been described. These techniques were built upon common image processing techniques, such as image thresholding, segmentation, and thinning. So far, these techniques have been tested and evaluated to inspect reinforced concrete cracks and spalling, and asphalt pavement cracks and potholes. The results with these machine vision techniques have shown that they can be used for the damage detection and measurement of reinforced concrete structures and the distress detection and measurement of asphalt pavements.

References

 1. Abdel-Qader I, Abudayyeh O, Kelly ME (2003) Analysis of edge-detection techniques for crack identification in bridges. J Comput Civ Eng 17(4):255–263
 2. Abdel-Qader I, Pashaie-Rad S, Abudayyeh O, Yehia S (2006) PCA-based algorithm for unsupervised bridge crack detection. Adv Eng Softw 37(12):771–778
 3. ACI 224.1R (2005) Causes, assessment and repair of cracks in concrete structures. ACI manual of concrete practice
 4. ACI 349.3R (2005) Evaluation of existing nuclear safety-related concrete structures. ACI manual of concrete practice
 5. Aldunate R, Ochoa SF, Peña-Mora F, Nussbaum M (2006) Robust mobile ad hoc space for collaboration to support disaster relief efforts involving critical physical infrastructure. J Comput Civ Eng 20(1):13–27
 6. Amarasiri S, Gunaratne M, Sarkar S (2010) Modeling of crack depths in digital images of concrete pavements using optical reflection properties. J Transp Eng 136(6):489–499
 7. Applied Technology Council (ATC) (1989) ATC-20, Procedures for Post Earthquake Safety Evaluations of Buildings. Report ATC-20, applied technology council, Redwood City, CA
 8. ASCE (2011) American society of civil engineering. National infrastructure report card. http://www.infrastructurereportcard.org. 10 Nov 2011
 9. Bartel J (2001) A picture of bridge health. NTIAC Nondestr Test Inf Anal Cent Newslett 27 (1):1–4
10. Bauer J, Sunderhauf N, Protzel P (2007) Comparing several implementations of two recently published feature detectors. In: Proceedings of the international conference on intelligent and autonomous systems, Toulouse, France
11. Bradski G, Kaehler A (2008) Learning OpenCV: computer vision with the OpenCV. O'Reilly. ISBN-13: 978-0596516130
12. Chae MJ, Iseley T, Abraham DM (2003) Computerized sewer pipe condition assessment. In: Proceedings of the ASCE international conference on pipeline engineering and construction Baltimore, MD, pp 477–493
13. Chang KT, Chang JR, Liu JK (2005) Detection of pavement distresses using 3D laser scanning technology. In: Proceedings of the 2005 ASCE international conference on computing in civil engineering, p 105
14. Chen PH, Chang PL (2006) Effectiveness of neuro-fuzzy recognition approach in evaluating steel bridge paint conditions. Can J Civ Eng 33:103–108
15. Chen PH, Yang YC, Chang LM (2010) Box-and-ellipse-based ANFIS for bridge coating assessment. J Comput Civ Eng ASCE 24(5):389–399
16. Chen PH, Shen HK, Lei CY, Chang LM (2012) Support-vector-machine-based method for automated steel bridge rust assessment. Autom Constr 23:9–19
17. Cheng H, Shi X, Glazier C (2003) Real-time image thresholding based on sample space reduction and interpolation approach. J Comput Civ Eng 17(4):264–272

18. Federal Emergency Management Agency (FEMA) (2006). National urban search and rescue response system—structure specialist position description. http://www.disasterengineer.org/library/Struct%20Spec%20PD%20July%202006.pdf. Dec 2008

19. Federal Highway Administration (FHWA) (1995) Recording and coding guide for the structure inventory and appraisal of the nation's bridges. US department of transportation, Washington, DC

20. Federal Highway Administration (FHWA) (2003) Distress identification manual for the long term pavement performance program. Publication number: FHWA-RD-03-031. http://www.tfhrc.gov/pavement/ltpp/reports/03031/03031.pdf

21. Federal Highway Administration (FHWA) (2009a) Questions and answers on the national bridge inspection standards 23 CFR 650 subpart C. US department of transportation. http://www.fhwa.dot.gov/Bridge/nbis/. 29 Sep 2009

22. Federal Highway Administration (FHWA) (2009b) LTPP—Beyond FY 2009: what needs to be done? Technical Report, Publication No. FHWA-HRT-09-052. http://www.fhwa.dot.gov/pavement/ltpp/pubs/09052/09052.pdf. Nov 2011

23. Fugro Roadware (2010) Application: pavement condition assessment, data sheets available at http://www.roadware.com/_lib/pdf/pavement_assesment.zip. Nov 2011

24. German Department of Transportation—Road and Transportation Research Association (FGSV) (2006) Additional technical contractual terms and guidelines for pavement condition survey and assessment (ZTV ZEB-StB), Report No. FGSV 998

25. German S, Brilakis I, DesRoches R (2012) Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments. Adv Eng Inform 26(4):846–858

26. Haas C (1996) Evolution of an automated crack sealer: a study in construction technology development. Autom Constr 4(4):293–305

27. Hou Z, Wang KCP, Gong W (2007) Experimentation of 3D pavement imaging through stereovision. In: Proceedings of international conference on transportation engineering (ICTE 2007), pp 376–381

28. Huang Y, Xu B (2006) Automatic inspection of pavement cracking distress. J Electron Imaging 15(1):013017

29. Iyer S, Sinha SK (2006) Segmentation of pipe for crack detection in buried sewers. Comput Aided Civ Infrastruct Eng 21:395–410

30. Jauregui D, White K (2003) Implementation of virtual reality in routine bridge inspection. J Transp Res Rec 1827:29–35

31. Jiaqiu W, Songlin M, Li J (2009) Research on automatic identification method of pavement sag deformation. In: Proceedings of the 9th international conference of Chinese transportation professionals (ICCTP), pp 2761–2766

32. Karuppuswamy J, Selvaraj V, Ganesh MM, Hall EL (2000). Detection and avoidance of simulated potholes in autonomous vehicle navigation in an unstructured environment. In: Casasent DP (ed), Intelligent robots and computer vision XIX: algorithms, techniques, and active vision, Proceedings of SPIE, vol 4197, pp 70–80

33. Kim YS, Yoo HS, Lee JH, Han SW (2009) Chronological development history of X-Y table based pavement crack sealers and research findings for practical use in the field. Autom Constr 18(5):513–524

34. Koch C, Brilakis I (2011) Pothole detection in asphalt pavement images. Adv Eng Inform 25(3):507–515

35. Koch C, Jog G, Brilakis I (2012) Automated pothole distress assessment using asphalt pavement video data. J Comput Civ Eng 27(4):370–378

36. Kostoulas D, Aldunate R, Peña-Mora F, Lakhera S (2006) A decentralized trust model to reduce information unreliability in complex disaster relief operations. In: Proceedings of the 13th EG-ICE workshop on intelligent computing in engineering and architecture, Ascona, Switzerland, LNCS 4200, Jun 25–30, pp 383–407

37. Koutsopoulos HN, Downey AB (1993) Privitive-based classification of pavement cracking images. J Transp Eng 19(3):136–143

38. Lee S, Chang LM, Skibniewski M (2006) Automated recognition of surface defects using digital color image processing. Autom Constr 16(4):540–549
39. Lee S (2010) An eigenvalue-based recognition method and its application to bridge coating. In: Proceedings of the 6th International Conference on Innovation in Architecture, Engineering, and Construction (AEC)
40. Li Q, Yao M, Yao X, Xu B (2010) A real-time 3D scanning system for pavement distortion inspection. Meas Sci Technol 21:8
41. Liu Z, Shahrel A, Ohashi T, Toshiaki E (2002) Tunnel crack detection and classification system based on image processing. Machine vision applications in industrial inspection X. Proc SPIE 4664:145–152
42. Minnesota Department of Transportation (MnDOT) (2003) Distress identification manual, 2003. http://www.dot.state.mn.us/materials/manuals/pvmtmgmt/distressmanual.pdf. May 2010
43. Minnesota Department of Transportation (MnDOT) (2009) 2009 pavement condition executive summary, Report Number: MnDOT/OMRR-PM–2009-01. www.dot.state.mn.us/materials/pvmtmgmtdocs/execsumm_2009.pdf. Nov 2011
44. Momma (2008) OpenCV thinning. Online source. http://www.eml.ele.cst.nihon-u.ac.jp/~momma/wiki/wiki.cgi/OpenCV/. 1 Mar 2011
45. Moore M, Phares B, Graybeal B, Rolander D, Washer GA (2001) Reliability of visual inspection for highway bridges, Volume I: Final report and, Volume II: Appendices US department of transportation, Washington, DC, FHWA-RD-01-020(021)
46. NASA (2009) Disaster assistance and rescue team—structural assessment team. http://dart.arc.nasa.gov/SAT/SAT.html. Dec 2009
47. Oh H, Garrick NW, Achenie LEK (1997) Segmentation algorithm using iterative clipping for processing noisy pavement images. In: Proceedings of the 2nd international conference on imaging technologies: techniques and applications in civil engineering, Switzerland, pp 138–147
48. Perenchio WF (1989) The condition survey. Concr Int 11(1):59–62
49. Phares B, Washer G, Rolander D, Graybeal B, Moore M (2004) Routine highway bridge inspection condition documentation accuracy and reliability. J Bridge Eng 9(4):403–413
50. Sinha S, Fieguth P (2006) Automated detection of cracks in buried concrete pipe images. Autom Constr 15(1):58–72
51. South Dakota Department of Transportation (SdDOT) (2009) Visual distress manual. http://www.sddot.com/pe/projdev/docs/distressmanual.pdf. Nov 2011
52. Sun Y, Salari E, Chou E (2009) Automated pavement distress detection using advanced image processing techniques. In: Proceedings of the international conference on electro/information technology, pp 373–377
53. Sunkpho J (2001) A framework for developing field inspection support systems. PhD Dissertation, Department of civil and environmental engineering, Carnegie Mellon University
54. Thompson P, Shepard R (2000) AASHTO commonly recognized bridge elements, successful applications and lessons learned national workshop on commonly recognized measures for maintenance. http://www.pdth.com/images/coreelem.pdf 21 Sep 2009
55. Tsai Y-C, Kaul V, Mersereau RM (2010) Critical assessment of pavement distress segmentation methods. J Transp Eng 136(1):11–19
56. Tsai Y, Jiang C, Huang Y(2012) A multi-scale crack fundamental element model for real-world pavement crack classification. J Comput Civ. Eng. doi:10.1061/(ASCE)CP.1943-5487.0000271. 15 Nov 2012
57. UK Department for Transportation (UKDFT) (2013). Technical note: road condition and maintenance data. https://www.gov.uk/transport-statistics-notes-and-guidance-road-conditions
58. Wang KCP (2004) Challenges and feasibility for comprehensive automated survey of pavement conditions. In: Proceedings of 8th international conference on applications of advanced technologies in transportation engineering
59. Wang KCP, Gong W (2005) Real-time automated survey system of pavement cracking in parallel environment. J Infrastruct Syst 11(3):154–164

60. Wang KCP (2007) Positioning and imaging sensors for automated asset management of transportation facilities. In: Proceedings of the 1st international conference on transportation engineering (ICTE)
61. Wasserman E (2007) Are Tennessee bridges safe? http://www.tdot.state.tn.us/documents/Wasserman_editorial.pdf 21 Sep 2009
62. Yamaguchi T, Hashimoto S (2009) Fast crack detection method for large-size concrete surface images using percolation-based image processing. Mach Vis Appl 11(5):797–809
63. Yu S-N, Jang J-H, Han C-S (2007) Auto inspection system using a mobile robot for detecting concrete cracks in a tunnel. Autom Constr 16(3):255–261
64. Zack GW, Rogers WE, Latt SA (1977) Automatic measurement of sister chromatid exchange frequency. J Histochem Cytochem 25(7):741–753
65. Zhu Z, German S, Brilakis I (2011) Visual retrieval of concrete crack properties for automated post-earthquake structural safety evaluation. Autom Constr 20(7):874–883

Chapter 12
Smart Check 3D: An Industrial Inspection System Combining 3D Vision with Automatic Planning of Inspection Viewpoints

Nicola Carlon, Nicolò Boscolo, Stefano Tonello and Emanuele Menegatti

Abstract In this chapter, we describe an industrial inspection system composed by a 3D vision system, mounted on a manipulator robot arm, able to perform quality and completeness inspection on a complex solid part. The novelty of the system is in the deep integration among three software modules: the visual inspection system, the 3D simulation software, and the motion planning engine of the manipulator robot. This enables an automatic off-line programming of the robot path by specifying in the system the desired inspection tasks. The system automatically generates the needed points of view in order to perform 3D reconstruction and automatic visual inspection. Moreover, the motion planning system can reorder the inspection points in order to optimize the inspection cycle time. The core of this system was developed in the European Project "Thermobot," and currently, it is been engineered to be deployed in an industrial production plant.

Contents

N. Carlon (✉) · N. Boscolo · S. Tonello
IT+Robotics Srl, Strada Prima 11, 35129 Padova, Italy
e-mail: nicola.carlon@it-robotics.it

N. Boscolo
e-mail: nicolo.boscolo@it-robotics.it

S. Tonello
e-mail: stefano.tonello@it-robotics.it

E. Menegatti
Intelligent Autonomous Systems Laboratory, University of Padova, via Gradenigo 6a, 35131 Padova, Italy
e-mail: emg@dei.unipd.it

12.1 Introduction

Quality control is a critical step in the manufacturing process, it aims to assess the quality standards of the product being manufactured, and it allows to detect defects in the earlier stages of production. The variety of products to be inspected is wide. Then, a new software design is needed to enable even the most unexperienced user to configure the quality checks to be carried out. Furthermore, if quality control is done online, it allows the operator to immediately take action in order to improve the process, for instance by replacing any parts held responsible for malfunctions, thus saving waste and optimizing the available resources. Most common industrial visual inspection systems rely on 2D-based image analysis [1] to perform completeness inspection. However, without scene depth information, 2D machine vision systems have various shortcomings that limit their capabilities and robustness [2].

By introducing laser triangulation methods in automatic visual inspection it is possible to capture the shape of the product and thus possible to compare it using a metric shape comparison approach. In this way, even small difference between shapes can be identified and evaluated. 3D vision technology can be combined with standard 2D shape acquisition, by using proper lighting to identify small defects which would otherwise be hard to spot by using only laser triangulation methods, such as for instance burrs inside holes, non-uniform coloring, or components of wrong color.

To increase flexibility, the vision system can be mounted on a robot. Alternatively, it is possible to pick the part to be inspected with the robot and to move the product in front of a fixed vision system. This introduces a further requirement to minimize the cycle time, making it mandatory, in addition to the use of vision algorithms, to generate collision-free path and optimize cycle time for a full inspection of the product from any angle. The narrow interaction between the visual inspection simulation and the motion planning enables the customer to save time on the design of the inspection cycle. Moreover, the choices computed by the

algorithms permit to save cycle time as well. Thanks to this kind of requirement and to the research carried out during European project 3DComplete (for visual inspection) and Thermobot (path and motion planning), a visual inspection system was created in order to allow a general purpose and a totally customizable tool. The resulting software, called Smart Check 3D, allows end-users to perform specific controls without necessarily requiring specific knowledge about programming or tuning vision algorithms.

The following paragraphs will explain how the state-of-the-art researches in motion planning, vision algorithms, and 3D reconstruction are implemented by a software that meets the current industry standards. The remainder of this chapter is organized as follows: in Sect. 12.2, the state of the art on machine vision and motion planning is presented. Then, in the Sect. 12.3, the main modules needed for the inspection workflow are presented. The modules are workcell definition, inspection device calibration, configuration of inspection tasks, generation of virtual reference object, task merging and reordering, motion planning and online execution. In Sect. 12.4, the presentation of a case study is given. Finally, conclusions and future works are drawn in Sect. 12.5.

12.2   State of the Art

12.2.1   Visual Inspection

Up to now, and more in the future, the industrial sector is facing a big change in order to achieve the lean production paradigm. A modern industrial plant has to offer a high level of configurability and range of applicability. The goal is to produce a large variety of fine product in a quantity that is related to a rapidly changing market response over the time. Many commercial solutions for visual inspection work only at a fixed resolution and sensing range or they need complex calibration tasks. Then, even the industrial software panorama needs new solutions.

Indeed, there actually exists a large number of solutions for completeness product inspection and part metrology (that use measuring machine based on vision or probes) [1], but almost all these solutions require a human operator or very long setup time to reach a complete inspection cycle. Moreover, unlike metrologic tests using touching probes, vision can be naturally parallelized, exploiting the parallelization trend [3].

Two of the main industrial application fields for the new frontiers of the machine vision are robot guidance and visual inspection for quality control. These two fields are both based on the object detection with comparison to a known model in order to detect position (e.g., random bin-picking [4]) or differences (e.g., part inspection). Unlike random bin-picking systems that always operate in real time, it is less usual to set up autonomous and highly reconfigurable industrial quality control systems in manufacturing cycle in order to control 100 % of the products.

Vision-based solutions can also speed up quality control tasks and can be easily deployed in the manufacturing cycle. Indeed, the natural way to perform part inspection with machine vision is 3D acquisition and processing. The 2D vision technology can be combined with 3D in particular tasks; sometimes, the use of the 3D acquisition can complement 2D vision when its limits are reached [5].

12.2.2   Motion Planning

The visual inspection often involves the robot manipulation (e.g., moving the vision system or the inspected product). The manipulation task has highly nonlinear dynamics, uncertain mechanics of contact, and under-actuated phases of object motion (e.g., flight), which pose difficulties for planning. These tasks require precise coordination in order to accelerate the robot to high speed to arrive at a particular configuration or set of configurations, while also remaining safe by avoiding collisions and joint limits. Safety is again a concern when decelerating to a stop. The motion planning can enhance the visual inspection performances from the trajectory quality side. Indeed, the trajectories used for the motion of the inspection devices (e.g., robot arms, UAV) can be optimized with the heuristics and the techniques exposed by the motion planners. Motion planning algorithms are exploited in a wide range of applications including industrial robotics, bioinformatics [6], and gaming/virtual reality [7]. The analytical approaches [8] rely heavily on analyzing the equations of motion and constraints on the system, which makes it difficult to handle new scenarios. Unfortunately, the general motion planning problem is known to be PSPACE-hard [9, 10]. In an effort to solve problems of practical interest, much recent research has been devoted to developing sampling-based planners such as the rapidly exploring randomized tree (RRT) [11], the probabilistic roadmap methods (PRM) [12], and ARW [13]. These algorithms only achieve probabilistic completeness. One of the key drawbacks of these sampling-based algorithms is that they often produce suboptimal solutions that wander through state space. In fact, Karaman and Frazzoli have shown that RRT almost surely does not find an optimal solution [14]. RRT requires the user to specify two domain-specific extension heuristics: a distance metric and a node extension procedure. Indeed, it has been shown that the performance of RRT-based algorithms is sensitive to the distance metric used [15]. Prior research has characterized the effect of the distance metric component on the convergence of RRT. LaValle and Kuffner have previously noted that cost-to-go functions could be based on, among other options, optimal control for linearized systems [11]. In [16] has been proposed an LQR-based heuristic to be used with the RRT algorithm similar to the one proposed by Glassman and Tedrake [17] for the RRT.

Learning approaches have been used for dynamic motion planning because previous random techniques outcomes are often sensitive to calibration errors and noise. Aboaf et al. [18] applied classical learning techniques to the problem of juggling a ball in three dimensions. Using several practice iterations, the learning

procedures adapt the task model and control policy to compensate for systematic inaccuracies. Such techniques require an initial policy that is sufficiently close to optimal in order to converge.

Another way to enhance the planning performances is to use the parallelized versions of the PRM algorithms. Plaku et al. [19] present SRT where there is a combination of sampling-based method primarily designed for multiple query motion planning (PRM) with sampling-based tree methods primarily designed for single query motion planning (expansive space trees, rapidly exploring random trees, and others) in a novel planning framework that can be efficiently parallelized. Results show that up to 24-DOFs they were able to obtain nearly linear speedup for parallel SRT.

12.3 Smart Check 3D

Smart Check 3D provides an integrated method for optimized off-line programming of quality control. The proposed system is fully customizable using a top-down approach: from CAD model to robot trajectory. Our solution aims to make easier the work of system integrators and to make easier the setup by the end-user of generic quality inspection tasks to increase his autonomy.

The workflow is simple: end-user defines his workspace, imports a CAD model of his product, and defines all inspection tasks that the system has to perform. The task design includes the selection of the inspection points and the needed quality controls over the CAD model (see 12.3.4). If user does not have the CAD model, he can also refer the controls to a reference acquisition and the configuration of defects can be postponed after the first execution. The definition of the inspection tasks does not require any knowledge of computer vision or image processing algorithms. The system provides a list of high-level inspection blocks such as "3D deformity," "measure," and "scratches." Each block exposes some customizable parameters that allow the user to give a quantitative definition of the defect to be detected. Virtual inspection can be implemented using different sensors. Smart Check 3D currently supports 2D camera and 3D laser triangulation system with one or two lasers.

The motion planning exploits the information regarding the product geometry and its inspection. Then, the planner will propose an optimized set of robot movements that perform the inspection tasks, minimizing cycle time. Using the robot vendor language, a robot program is automatically generated, and it can be directly loaded and executed. The following subsections describe the modules involved in the workflow steps:

- entities definition (workcell, inspection devices, reference product);
- Setup of inspection tasks; and
- Generation of robot code vendor-specific and online cycle (Fig. 12.1).
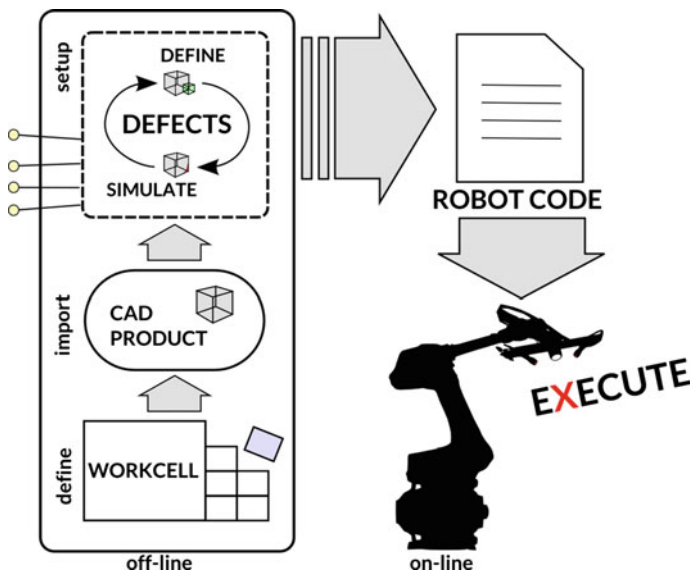
Fig. 12.1 Smart Check 3D workflow phases: cell definition, 3D model import, and the setup of the inspection routine

12.3.1 Workcell Definition

The first step is the definition of the inspection workcell that has to be simulated in a virtual 3D framework. This is accomplished using Workcell Simulator [20], a 3D simulator to create virtual industrial environments in order to test and simulate industrial processes (e.g., metal-sheet bending, bin-picking, visual inspection). Workcell Simulator allows to accurately position all components using standard calibration procedures and to simulate the motion of the robots and of other components (Fig. 12.2).

The definition of the virtual workcell environment includes also 2D and 3D vision sensors to support virtual sensor acquisition. By doing this, we can test vision algorithms to identify differences between the correct model provided by the user and a defective model auto-generated by the system and based on the expected defects.

12.3.2 Inspection Device Calibration

To allow correct simulation and visual data acquisition of the inspection process, the workcell needs to be calibrated.

Calibration procedure is divided into three steps:

Fig. 12.2 Screenshot of inspection workcell used in the European Project Thermobot showing the autonomous robotic system for the thermo-graphic detection of cracks
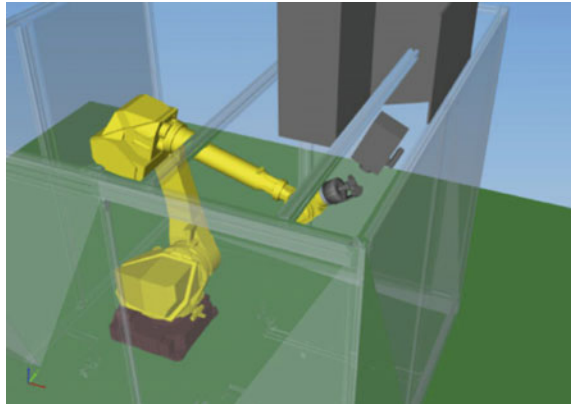
1. Camera intrinsic parameters (focal length, principal point, distortion): this is accomplished using OpenCV calibration procedure [21];
2. Rigid transformation from camera reference frame to robot reference frame: this is accomplished using Tsai's method [22];
3. Plane laser calibration: this is accomplished using methods described in [23].

All these calibrations do not take long time because consist in acquiring a small set of images of standard checkerboards in different poses. Because of inspection workcell worsening, the physics of the environment could change over time. Then, the calibration parameters may change as well with a precision loss. Once the calibration task is defined, it can be performed after a fixed term cyclically, in order to ensure maximum system performance over time. Our vision calibration and configuration system are based on the results of the "3DComplete" European project where two lasers and a peak detection camera are involved. The two lasers are used to reduce occlusions in order to obtain a complete item inspection. The 3D vision system is one of the fully configurable sensors, allowing advanced features such as positioning the two lasers asymmetrically to vary sensing range and depth resolution.

12.3.3 Configuration of Inspection Tasks

Once the calibration is completed, the inspection tasks can be designed. As first step, the correct model of the inspected part is loaded inside the simulator using a CAD model or with the 3D scanning system from a real part without defects. Then, the user can configure the inspection cycle, as a set of inspection tasks. The task is a set of three elements: the acquisition over the product, the related quality inspection controls, and the robot trajectory. The core of the task is based on a model buildup from the acquisition process that basically answers the question "which part of the surface can be inspected from a given viewpoint?" Moreover, the user can specify a

set of 3D region of interest, customizing dimensions, and positions for 3D bounding box in which to search for defects. For each defect, the user can define dimensional tolerances. When the task is built, in order to check whether it is feasible, the configuration module is integrated with kinematic and dynamic models of the robots to be used for the task, and it is extended with motion planning to avoid collisions (Fig. 12.3).

12.3.4 Generation of Virtual Reference Object

Our system uses ray tracing to automatically generate a virtual 3D scan starting from the model, according to the resolution of the laser triangulation system and all its parameters given by the calibration step (see Fig. 12.4).

This simulation helps and gives an instant feedback to the user about the quality of the parameters used in the detection. This is done by executing off-line quality control tasks and virtual defect generation. In order to avoid disruptive tests, the user can generate virtual defects from specifications provided by the user such as position, size, and granularity. 2D detectable defects can be defined as well. The system provides the point of view needed to perform the correct acquisition of defect and visualize a virtual, but realistic image of the acquisition. The workflow is very simple, when the virtual reference object is ready (see Fig. 12.5a), a 3D tool enables the user to remove in a fine way the set of points from the cloud (see Fig. 12.5b). Then, a 3D box representing the region of interest can be placed over the point cloud (see Fig. 12.5c). Finally, the off-line quality control can be performed (see Fig. 12.5d).
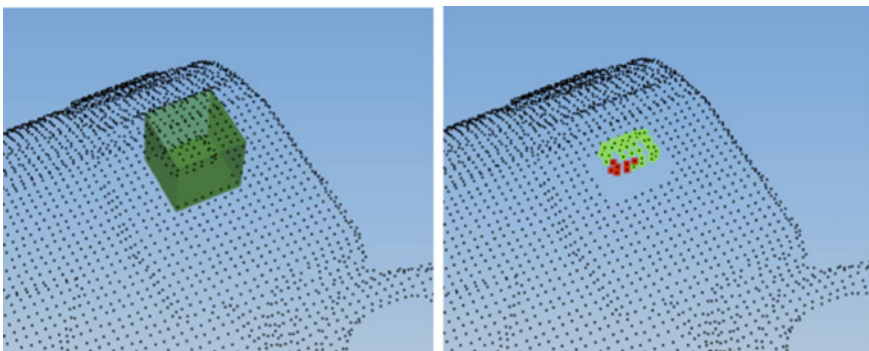
Fig. 12.3 Definition of a 3D region of interest in order to test the inspection quality using virtual defects testing
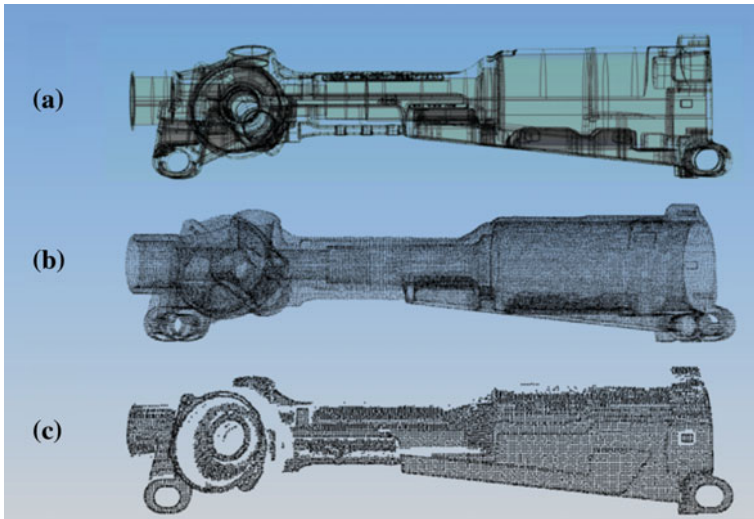
Fig. 12.4 Model view generation: **a** CAD drawing STP; **b** Model voxelization **c** Synthetic view generated by the sensor acquisitions
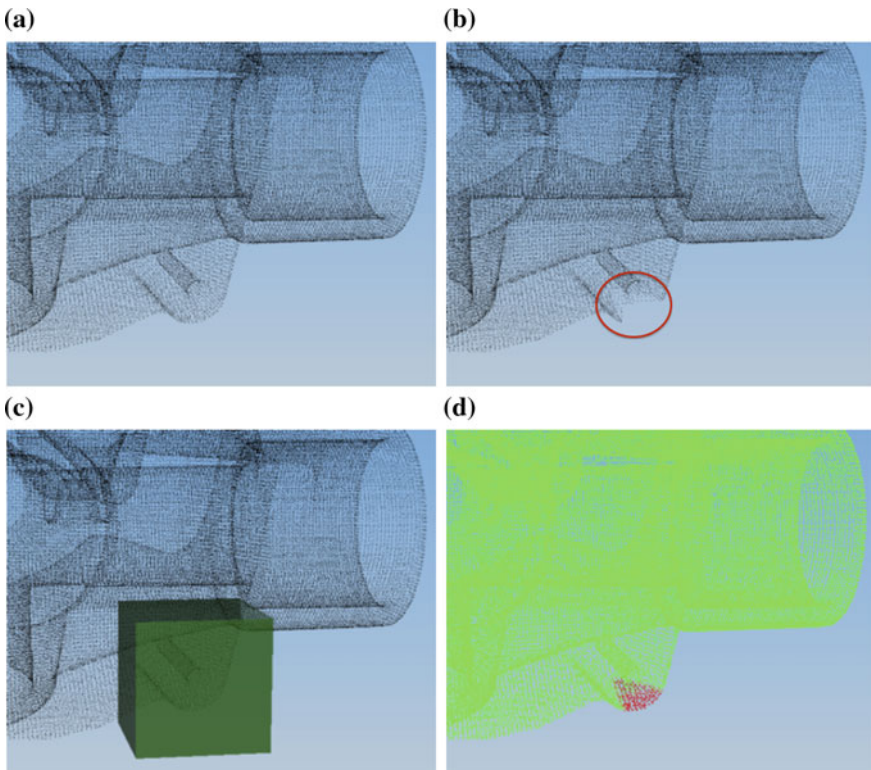
Fig. 12.5 The phases of visual inspection testing: from the virtual defect generation to the recognition. **a** Original model view. **b** Generated defect. **c** User defined region of interest. **d** Result of off-line execution of quality control

12.3.5 Task Merging and Reordering

When a high number of inspection points are defined, the inspection cycle might significantly increase in time. Often, it is possible to improve the scanning time by carefully analyzing the inspection sequence and the path between the inspection points. Indeed, in most cases it is possible to calculate the optimal acquisition trajectory that includes all inspection point and that minimizes the scanning time. The definition of a group of different neighboring defects may cause performance degradation in terms of cycle time. Task merging consists in finding a unique acquisition. This is possible thanks to the sensor simulation that considers occlusions and selects the correct trajectory, minimizing robot movements and stops between acquisitions. We called this process "task merging." Moreover, on top of task merging optimization, another optimization can be done in order to reduce cycle time: what we called "task reordering." The task reordering problem showed up for the first time during the Thermobot project [24] when we had different part of a product to inspect. The position of these tasks and the order used to perform the inspection can produce redundant robot paths increasing the inspection cycle time. In order to resolve this issue, we have designed a complete graph where the nodes represent the tasks and the edges are weighted with the time spent by the robot to drive through the two neighboring tasks. This problem is well known in the literature as the Asymmetric TSP[1] where the function to optimize is the time sequence. The ATSP problem aims to find the minimum path that covers all the nodes. Minimum path among two nodes means the path that minimizes the sum of the cost along the edges contained in the path. Unfortunately, the problem is NP-complete so in order to find a suboptimal solution, algorithms based on ant colony optimization have been developed [25].

12.3.6 Motion Planning

The aim of the off-line motion planning module found the right robot arm joint values in order to avoid collisions between one inspection position and the following one. The planning and the optimization are performed over a configuration space of the robot kinematic (joint positions) and dynamic variables (joints velocity and acceleration) in order to plan a motion. A valid configuration is called keypoint, and it corresponds to a collision-free placement of the robot in its physical workspace. The temporal sequence of these keypoints is called trajectory. The planner will add other configurations among each couple of keypoints with the aim of avoid collisions. The collision checking between two configurations is performed using a linear interpolation of the robot motions. Our MP module is using a planning architecture instead of just one algorithm. The architecture is using several planning

---

[1]travel salesman problem.
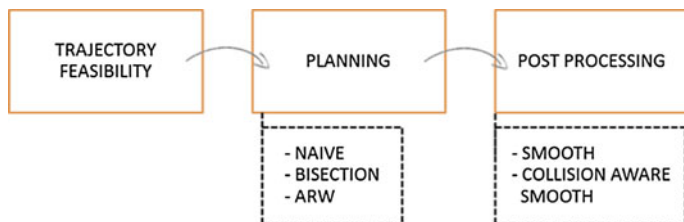
Fig. 12.6 Motion planning architecture overview: the modular architecture helps the planning core to find good solutions spanning a wide area of environments

algorithms divided into three sequential steps as shown in Fig. 12.6. Each step will use its own algorithms from the faster and less accurate to the slower and more precise. If one algorithm inside the test will fail, the architecture will select the next one inside the step. Each planner performs a backward configuration search: from the last keypoint to the first one.

In the first step, the feasibility of the trajectory keypoints is checked. A keypoint is feasible if it corresponds to a configuration reachable by the robot and it is not colliding. If the first step ends in a good way, the motion planners are executed in a hierarchical order, from the most simple and fast to the most effective but slower. A planner is executed only if the previous one fails to find a solution following this order:

1. *Naive*. Given the configurations sequence, the algorithm is going to check whether each couple of points is connectible[2]; otherwise, it will try to add a point between them. Usually, the middle point is colliding or the three points are not connectible, and then, the algorithm tries to shift this point editing each single-point dimension using some correction factors.
2. *Bisection*. This algorithm uses a logic that is similar to the previous one with the following two main different features:

   (a) A middle point is moved using perturbation random points over a normal distribution.
   (b) When a collision-free middle point is found, the connectible test is performed. But in this case, each segment (first-middle and middle-second keypoints) is treated recursively using a divide-and-conquer technique.

3. *ARW*. The Adaptive Random Walks algorithm is part of the probabilistic roadmap methods. The proposed algorithm turns out to be simple to implement, and the solution it produces can be easily and efficiently optimized. Furthermore, the algorithm can incorporate adaptive components in order to be

---

[2]Two keypoints are connectible if the robot motions needed to reach the second point starting from the first one are collision-free.
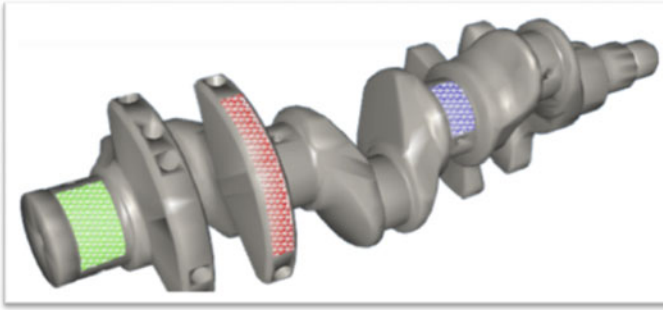
Fig. 12.7 Zones merging over a mechanical engine part

adapted to particular problems. Proofs of the theoretical soundness of the algorithm are provided in [13].

After the kinematic evaluation, the optimization algorithms will edit the key-points in order to make a smoother robot trajectory (Fig. 12.7).

12.3.7 Online Execution

When the off-line simulation and inspection setup is done and the robot code is uploaded to the robot controller, the online execution can be performed. During the online inspection phase, the system is mainly composed by four components: PLC (industrial computer), the inspection software, the robot, and inspection tool (camera). In the early stage, an external signal notifies to the PLC that a new inspection cycle should be performed. Then, the PLC sets up the robot and tells to inspection software to start the initialization of the inspection procedure. When the initialization is completed, the software ends the setup phase sending an acknowledge message to the PLC. Now, the PLC triggers the camera and the robot in order to start the frame acquisitions and robot movements. As long as the acquisition stream is performed, the software, when enough acquisitions are available, starts the defects checking using the internal algorithms. When the robot inspection trajectory is finished, the PLC sends an "end notice" message to the software and asks to the software whether the part is good or not. Finally, Smart Check 3D gives a response to the PLC in order to move the inspected product in the good or bad products box. It is not so unusual that some particular inspection cycle is slightly different from the normal one. For example, for a particular product, highly customizable inspections are needed. In this case, the sequence of trajectories can be partially replanned (without do everything from scratch) adding or removing the new task needs (Fig. 12.8).
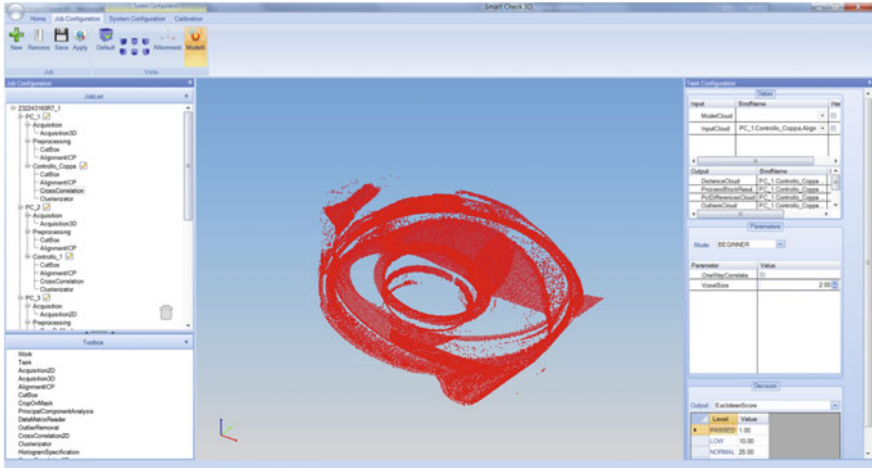
Fig. 12.8 Smart Check 3D screenshot during a task definition

12.4   Case Study

A significant use case of Smart Check 3D was the inspection of cast aluminum parts for a leading supplier of systems for the European automotive industry. The system is composed by the following:

- a robot equipped with a smart camera with laser peak detection hardware mounted on the TCP
- illuminator for 2D defects and DataMatrix recognition
- laser triangulation system for 3D inspection (Fig. 12.9)
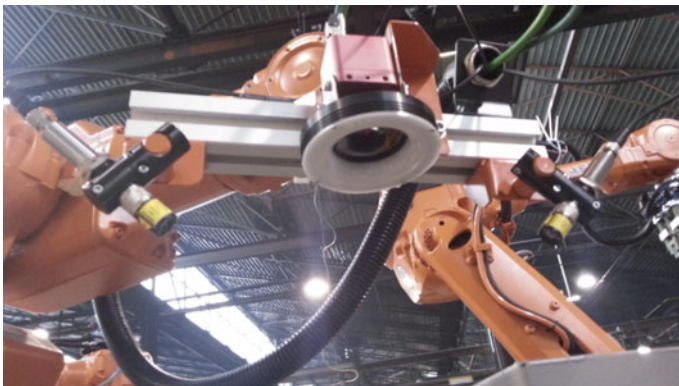
Fig. 12.9 Smart Check 3D screenshot during a task definition
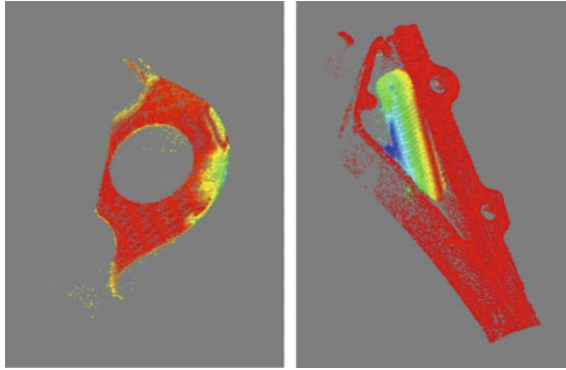
Fig. 12.10 3D inspection: completeness task
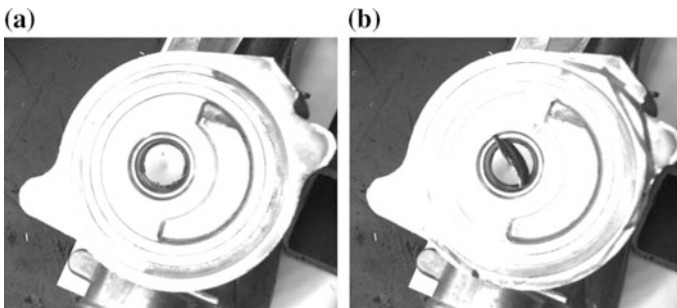
Fig. 12.11 2D visual inspection: inclusion checking. **a** Reference 2D Image. **b** Slave image

As aforementioned, Smart Check 3D has been used even for 2D and 3D visual inspection. In this case of study, different sequences of tasks have been set up and the tasks sequence replanning performed as well. The replan is needed during some critical moments of the production process, where the usury could influence some products in a particular ways. For those products, some tasks inside a sequence needs to be added or changed. With reference to the 3D inspection, a completeness task (see Fig. 12.10) has been performed over many areas of the product using the laser technology mentioned before. From the 2D inspection side, an inclusion task was requested (see Fig. 12.11).

12.5 Conclusions and Future Works

In this chapter, we described an industrial inspection system composed by a 3D vision system able to perform quality and completeness inspection on a complex solid part. The novelty of the system is in the deep integration among three software

modules: the visual inspection system, the 3D simulation software, and the motion planning engine of the manipulator robot. The system plans the needed point of view in order to perform the 3D reconstruction and inspection and can reorder the inspection points in order to optimize the inspection cycle time. The core of this system was developed in the European project 3DComplete (for visual inspection) and Thermobot (path and motion planning). Currently, it has been engineered to be deployed in an industrial production plant. We plan to extend the software including metrology tasks. In order to introduce more flexibility to the inspection sequence, we are considering to introduce online trajectory replanning. This feature can be useful when it is not clear how many times a task has to be performed in order to declare whether a product is good or not. This case could happen when the inspection task response is sensible to the product physical features and multiple trials are needed to proof the response task reliability.

References

1. Malamas EN, Petrakis EGM, Zervakis M, Petit L, Legat JD (2003) A survey on industrial vision systems, applications and tools. Image Vis Comput 21(2):171–188
2. So EWY, Munaro M, Michieletto S, Tonello S, Menegatti E (2013) 3Dcomplete: efficient completeness inspection using a 2.5D color scanner. Comput Ind 64(9):1237–1252
3. Keyes D (2007) Petaflop/s, seriously. Lect Notes Comput Sci 4873:2
4. Pretto A, Tonello S, Menegatti E (2013) Flexible 3D localization of planar objects for industrial bin-picking with monocamera vision system. In: IEEE international conference on automation science and engineering (CASE), 2013, pp 168–175
5. So E, Munaro M, Michieletto S, Menegatti E, Tonello S (2013) 3Dcomplete: efficient completeness inspection using a 2.5D color scanner. Comput Ind Spec Issue 3D Imaging Ind 64(9):1237–1252
6. Singh AP, Latombe JC, Brutlag DL (1999) A motion planning approach to flexible ligand binding. In: ISMB, pp 252–261
7. Lien JM, Bayazit B, Sowell RT, Rodriguez S, Amato NM (2004) Shepherding behaviors. In: Proceedings. ICRA'04. IEEE international conference on robotics and automation, 2004, vol 4, pp 4159–4164
8. Mettin U, Shiriaev AS, Freidovich LB, Sampei M (2010) Optimal ball pitching with an under actuated model of a human arm. In: IEEE international conference on robotics and automation (ICRA), 2010, pp 5009–5014
9. LaValle SM (2006) Planning algorithms. Cambridge University Press, Cambridge
10. Schwartz JT, Sharir M (1983) On the piano movers problem. II. General techniques for computing topological properties of real algebraic manifolds. Adv Appl Math 4(3):298–351
11. LaValle SM, Kuffner JJ (2001) Randomized kinodynamic planning. Int J Robot Res 20 (5):378–400
12. Kavraki LE, Svestka P, Latombe JC, Overmars MH (1996) Probabilistic roadmaps for path planning in high-dimensional configuration spaces. IEEE Trans Robot Autom 12(4):566–580
13. Carpin S, Pillonetto G (2005) Motion planning using adaptive random walks. IEEE Trans Robot 21(1):129–136

14. Karaman S, Frazzoli E (2011) Sampling-based algorithms for optimal motion planning. Int J Robot Res 30(7):846–894
15. SM LaValle (2003) From dynamic programming to RRTs: algorithmic design of feasible trajectories. In: Control problems in robotics. Springer, Berlin, pp 19–37
16. Perez A, Platt R, Konidaris G, Kaelbling L, Lozano-Perez T (2012) LQR-RRT*: optimal sampling-based motion planning with automatically derived extension heuristics. In: IEEE international conference on robotics and automation (ICRA), 2012, pp 2537–2542
17. Glassman E, Tedrake R (2010) A quadratic regulator-based heuristic for rapidly exploring state space. In: IEEE international conference on robotics and automation (ICRA), 2010, pp 5021–5028
18. Aboaf EW, Drucker S, Atkeson CG (1989) Task-level robot learning: juggling a tennis ball more accurately. In: IEEE international conference on robotics and automation, 1989. Proceedings 1989, pp 1290–1295
19. Plaku E, Bekris KE, Chen BY, Ladd AM, Kavraki LE (2005) Sampling-based roadmap of trees for parallel motion planning. IEEE Trans Robot 21(4):597–608
20. Tonello S, Zanetti GP, Finotto M, Bortoletto R, Tosello E, Menegatti E (2012) Workcellsimulator: a 3D simulator for intelligent manufacturing. In: Simulation, modeling, and programming for autonomous robots. Springer, Berlin, pp 311–322
21. Bouguet J-Y, Perona P (1998) Camera calibration from points and lines in dual-space geometry. In: Proceedings of 5th European conference on computer vision, pp 2–6
22. Tsai RY (1987) A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. IEEE J Robot Autom 3(4):323–344
23. So EWY, Michieletto S, Menegatti E (2012) Calibration of a dual-laser triangulation system for assembly line completeness inspection. In: IEEE international symposium on robotic and sensors environments, pp 138–143
24. Eitzinger C, Ghidoni S, Menegatti E (2013) Thermobot: towards semi-autonomous, thermographic detection of cracks. In: International conference on heating by electromagnetic sources, Padua
25. Dorigo M, Maniezzo V, Colorni A (1996) Ant system: optimization by a colony of cooperating agents. IEEE Trans Syst Man, Cybern Part B Cybern 26(1):29–41

Chapter 13
Ultrasonic Evaluation and Imaging

**Susan L. Crawford, Michael T. Anderson, Aaron A. Diaz,
Michael R. Larche, Matthew S. Prowant and Anthony D. Cinson**

Abstract Ultrasonic evaluation of materials for material characterization and flaw detection is as simple as manually moving a single-element probe across a specimen and looking at an oscilloscope display in real time or as complex as automatically (under computer control) scanning a phased-array probe across a specimen and collecting encoded data for immediate or off-line data analyses. The reliability of the results in the second technique is greatly increased because of a higher density of measurements per scanned area and measurements that can be more precisely related to the specimen geometry. This chapter will briefly discuss applications of the collection of spatially encoded data and focus primarily on the off-line analyses in the form of data imaging. Pacific Northwest National Laboratory (PNNL) has been involved with assessing and advancing the reliability of inservice inspections of nuclear power plant components for over 35 years. Modern ultrasonic imaging techniques such as the synthetic aperture focusing technique (SAFT), phased-array (PA) technology, and sound field mapping have undergone considerable improvements to effectively assess and better understand material constraints.

Contents

S.L. Crawford (✉) · M.T. Anderson · A.A. Diaz · M.R. Larche · M.S. Prowant
Pacific Northwest National Laboratory, PNNL, Richland, WA, USA
e-mail: Susan.Crawford@pnnl.gov

A.D. Cinson
Nuclear Regulatory Commission NRC (Formerly PNNL), Washington D.C., USA

13.1 Introduction

An ultrasonic evaluation of a specimen, for example, a welded component in a nuclear power plant, can be used for a volumetric assessment of component integrity, thus helping to ensure the safe operation of the plant. This chapter will discuss the basic principles of ultrasonic inspections and the challenges in evaluating coarse-grained materials that are prevalent in the primary coolant piping systems in pressurized water reactors. The use of non-encoded data, where data points are typically manually acquired and there is no correlation of data and position on the specimen, is discussed. The merits of encoded data, usually acquired with an automated scanner, where each data point is associated with a scan or scan and index position, are presented. Examples of both non-encoded and encoded data are presented. Ultrasonic imaging principles are discussed as well as off-line signal processing with the synthetic aperture focusing technique. Finally, phased-array data acquisition and imaging principles are presented with applications in stainless steel components.

13.2 Principles of Ultrasonics

Ultrasonic evaluation of materials involves the propagation of mechanical waves in the material under test. The mechanical waves are introduced into the material via a piezoelectric probe or transducer that converts an electrical excitation into a mechanical wave. The return signal is captured with either the same probe, operating in a pulse-echo arrangement, or with a separate receive probe, commonly operating in a pitch-catch mode. The receive probe converts the mechanical wave into an electronic signal that is viewed on an oscilloscope or digitally recorded for online and off-line viewing. While ultrasonic energy is above the audible frequency range of 20 Hz–20 kHz, the typical transducer operates in the 0.1–50 MHz range but can run into the gigahertz range as with an acoustic microscope system.

Ultrasonic or acoustic wave applications are diverse and include such areas as characterizing materials in terms of elastic properties or for flaw detection for product acceptance or for assessing product integrity or remaining useful life. This later application is widely used in the energy sector with the focus of this chapter in assessing nuclear power plant component integrity.

The principle behind flaw detection relies on insonifying the region of interest in a component and receiving a reflection from the flaw. Reflections occur at an interface with an acoustic impedance mismatch between the two materials. The acoustic impedance is defined as follows [1]:

$$Z = \rho V \tag{13.1}$$

where Z = acoustic impedance, ρ = material density, and V = material velocity.

A large difference in impedance between two media results in a large reflection. An air-filled crack in a stainless steel pipe would be an example of a large impedance difference. Depending on the composition of the specimen, the specimen geometry, and the ultrasonic probe frequency, the flaw may be readily detected or go undetected.

The selection of probe frequency for a particular evaluation depends on the material composition and thickness. As acoustic waves travel through a material, they are attenuated. This attenuation is attributed to either absorption, conversion of acoustic energy to heat, or scattering. Scattering occurs in an isotropic material with grains or inhomogeneities having acoustic impedance values that are different than the surrounding material, and grain diameters are small or comparable to the acoustic wavelength. Additionally, scattering increases with frequency. Typically, lower frequencies are used for the evaluation of thicker specimens and those that are more attenuating or contain large grains. However, lower frequencies result in lower image resolution. Therefore, the probe frequency must be carefully selected for a particular application to balance resolution and adequate insonification of the area of interest. Typically, fine-grained steel specimens can be effectively evaluated with probes in the 2.25–5 MHz range, depending on specimen thickness. Probes with center frequencies in the 0.8–2.0 MHz range can be used on coarse-grained specimens with thicknesses less than 5 cm (2 in.). Frequencies in the 0.2–1.0 MHz range are typically used in specimens above 5 cm (2 in.) thick with 0.5-MHz probes producing a good balance between resolution and penetration.

A challenging material for ultrasonic evaluation is cast stainless steel (CSS). This material has been used extensively in the primary coolant piping systems of Westinghouse-designed pressurized water reactors (PWRs). Alloying elements and casting processes used in the fabrication of CSS materials are responsible for its corrosion resistance and strength but also create complex and coarse-grained microstructures. CSS piping is subjected to periodic volumetric examination based on requirements found in the American Society of Mechanical Engineers (ASME) Boiler and Pressure Vessel Code, Section XI, Rules for Inservice Inspection of Nuclear Power Plant Components. Ultrasonic testing (UT) is the standard method for inspecting the cast austenitic stainless steel welds in the primary piping. Centrifugally cast stainless steel is an anisotropic and inhomogeneous material. The manufacturing process can result in the formation of long columnar (dendritic) grain structures (approximately normal to the surface), with grain growth oriented along the direction of heat dissipation, often several centimeters in length. Additionally, during the cooling and solidification process, columnar, equiaxed (randomly

Fig. 13.1 Example of columnar microstructure in CSS piping on *top* and equiaxed microstructure on *bottom*
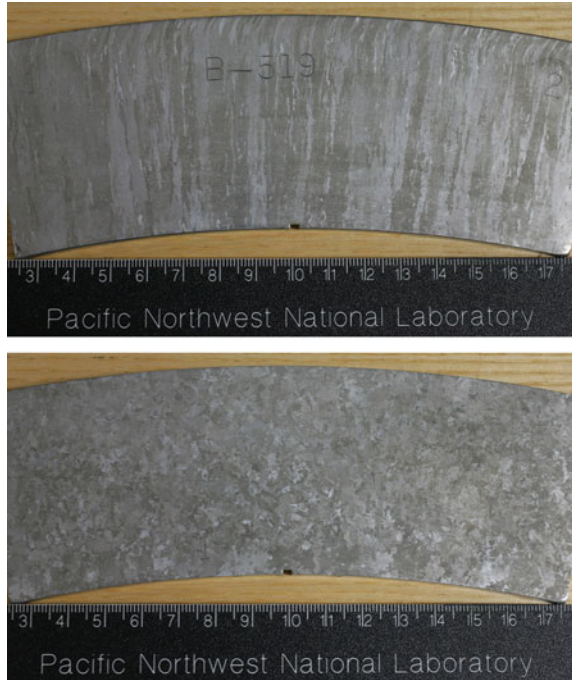
speckled microstructure), or a mixed structure can result, depending on chemical content, control of the cooling, and other variables in the casting process [2]. Examples of CSS piping microstructure are presented in Figs. 13.1 and 13.2 [3].

The large size of the anisotropic grains, relative to the acoustic pulse wavelength, strongly affects the propagation of ultrasound by causing severe attenuation, changes in velocity, and scattering of ultrasonic energy. Refraction and reflection of the sound beam can occur at the grain boundaries, resulting in defects being incorrectly reported, specific volumes of material not being examined, or both. When piping components are inspected from the outside surface, where the returning signal-to-noise ratio (SNR) is relatively low, ultrasonic examination can be confusing, unpredictable, and unreliable [4]. To reduce the impact of the microstructure on the ultrasonic inspection technique, frequencies in the 400 kHz–1.0 MHz range are used [2, 5].

13.3   Non-encoded Manual Scanning

During a manual scanning evaluation, the operator moves the transducer on a component while simultaneously watching a real-time display of the returned signal for a flaw response. A flaw is typically identified by a response that exceeds the

Fig. 13.2 Examples of mixed and banded microstructures in CSS piping

background noise level or some predetermined threshold in the area of interest. Figure 13.3 shows an example of the output signal from a portable flaw detector. This rectified A-scan signal (amplitude vs. time) represents an amplitude or voltage in the vertical direction and time or distance into a component in the horizontal direction. A threshold gate marked by the horizontal line was defined over some distance in the material, typically representing the area of interest in the component, and at some defined voltage level. The arrow points to a signal that exceeded the threshold, was in the gated time region and is therefore considered a flaw response.

Advantages to manual scanning are that a minimal amount of equipment is needed and the process can be quickly completed, resulting in an inexpensive examination. A probe and portable flaw detector with pulser–receiver electronics and a signal display is all that is required. This eliminates the need for a motorized scanner, controller, and an off-line data analyses platform, as is required for an encoded phased-array system.

Disadvantages to manual scanning are numerous. When there is no archived record of the data, flaw detection and characterization analyses must be performed in real time. Without a record of the data, the possibility of a data review does not

Fig. 13.3 Ultrasonic signal
from a manual evaluation
with *arrow* pointing to a
signal exceeding the threshold
(*horizontal line*)
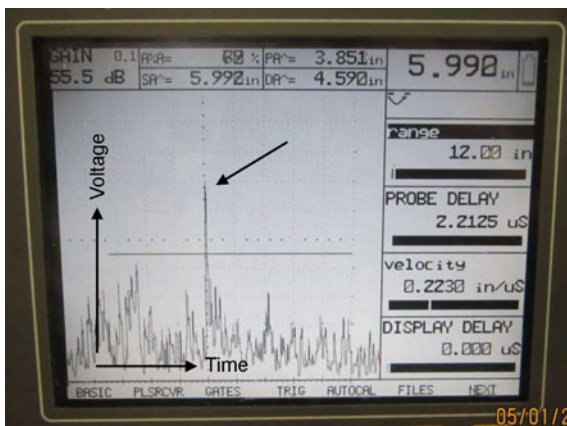
exist. Flaw detection and further characterization must be based on an A-scan signal
alone and do not allow for image-based reviews and analyses. If the operator does
not thoroughly cover the region of interest or if the flaw response very briefly
appears on the real-time A-scan display, the flaw could go undetected. As an
example, a dissimilar metal weld with nominal wall thickness of 122 mm (4.80 in.)
was examined with a 1-MHz probe. Assuming a scanning rate of 50.8 mm/s
(2.0 in./s), the response from a 33 % through-wall deep flaw was visible for less
than one second [6]. With such a low persistence, this flaw could easily go
undetected. Even with the archived data, there still is no precise correspondence of
the data to specimen geometry leading to possible misclassification of responses
and incorrect sizing.

13.4 Encoded Scanning

The use of encoders while collecting data provides spatial information for each
A-scan signal. Data are typically acquired while scanning in a line, Fig. 13.4, or
while raster scanning, Fig. 13.5. In each of these schematics, the green lines
emanating from the bottom of the wedge in the probe assembly represent the
generated sound field. This diagramed sound field represents a sweep of refracted
angles as could be generated with a phased-array (PA) probe. This could also be a
single refracted angle as generated from a conventional monolithic probe. An
example of a simple encoder wheel attached to a pitch-catch transducer configu-
ration is presented in Fig. 13.6 and is useful for acquiring data in a line as the
assembly is moved circumferentially around a pipe. In this example, time-of-flight
diffraction (TOFD) data are collected on a fusion joint in a high-density polyeth-
ylene (HDPE) pipe. The two monolithic probes are operating in a pitch-catch mode
with one probe transmitting and the other receiving. The fusion joint is centered

Fig. 13.4 Line scan data
acquisition on a welded
specimen with weld center
line (WCL) noted
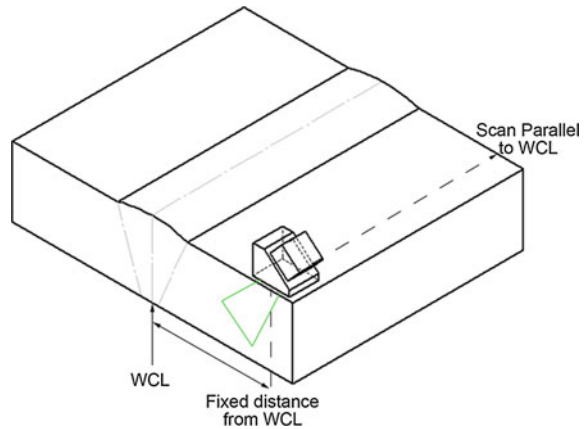
Fig. 13.5 Raster scan data
acquisition
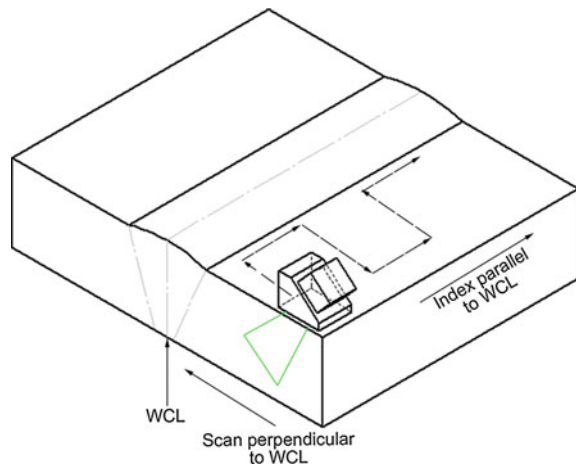
between the two probes. Acquired images are presented in Fig. 13.7 and represent left to right, responses from 75, 50, 25, and 10 % through-wall deep saw-cut notches [7], which were machined into the fusion joint for ultrasonic calibration purposes. The pipe outer diameter (OD) and inner diameter (ID) references are given in the figure, and red arrows mark the notch responses.

Consider raster scanning as depicted in Fig. 13.5 for a given pipe section containing an ID-connected circumferentially oriented flaw in the weld or adjacent material (heat-affected zone, HAZ). Crack initiation typically occurs in this location because of residual stresses from the welding process, the plant operating environment (water chemistry and temperature), and pipe and weld material. Raster scan lines are acquired perpendicular to the weld and can be collected in a bi- or unidirectional mode. Scanner systems typically have mechanical backlash at the end and start of the next scan line. While bidirectional data acquisition is faster, the
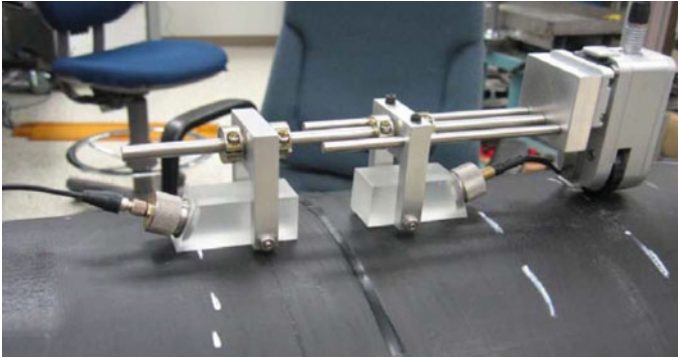
Fig. 13.6 Example of line scan data acquisition with an encoder wheel for the scan axis
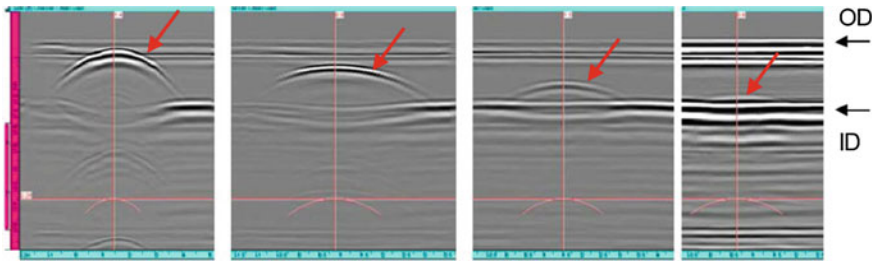
Fig. 13.7 TOFD data acquired with the setup from Fig. 13.6; *left* to *right* 75, 50, 25, and 10 % through-wall deep saw-cut notch responses

odd and even data lines are not properly registered. Unidirectional scanning eliminates this problem and is the preferred method unless the data lines are aligned with a post-processing algorithm. The data volume that is acquired with either a single-element or phased-array probe is typically viewed for off-line data analyses in three orthogonal planes. By convention, the B-scan or side view is an observation of a plane of the data volume with the horizontal representing the scan axis and the vertical representing the specimen thickness (T) or through-wall extent (Fig. 13.8). The B-scan could represent a single index plane, a selected range of index planes, or all index planes in the scan volume. If more than one index plane is selected, the B-scan image is a composite view of all the planes. More specifically, the maximum response at each scan and depth position from all selected index planes are displayed. The D-scan end view corresponds to a plane along the index axis. Therefore, the image horizontal represents the index axis, and the vertical represents the specimen thickness. Finally, a C-scan top view is a bird's eye view of the data volume with horizontal and vertical axes representing scan and index directions.

Encoded scanning with an automated scanner system greatly reduces acquisition time and provides greater precision in positioning, thus improving image resolution
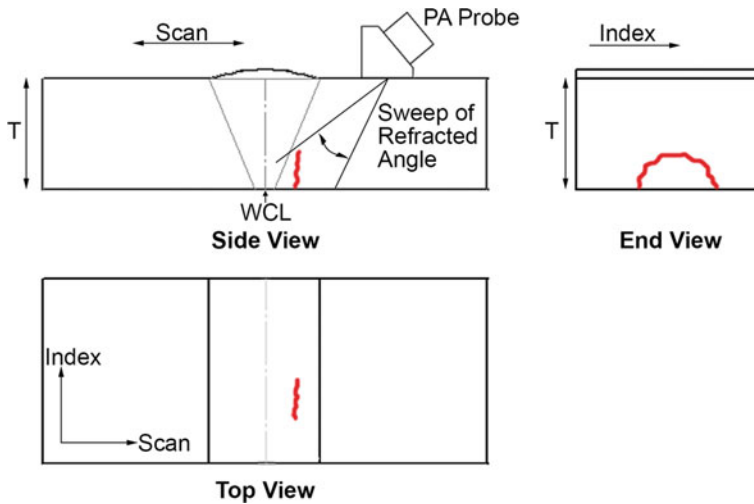
Fig. 13.8 Common conventional data presentations with a flaw noted in *red* (color figure online)

Fig. 13.9 Automated scanner and probe system on an elbow specimen
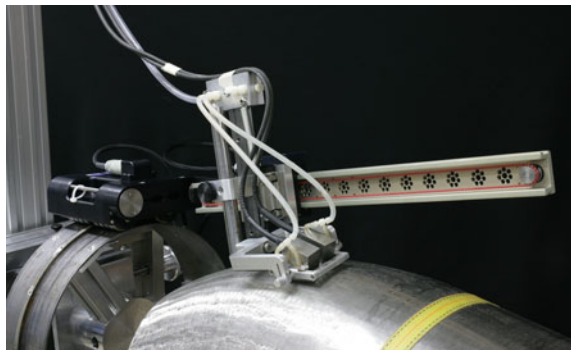
due to the smaller realized increments between scan and index points. An example of an automated scanner is displayed in Fig. 13.9 [5] and has a magnetic wheel assembly following a ring track (typically for the index axis) and a scanner arm that is commonly used for the scan axis.

13.5 Applications of Encoded Data

Several applications of encoded raster data collection and analysis are presented. The first example involves data collection with a monolithic probe and off-line data processing with the synthetic aperture focusing technique (SAFT). The second area

presented uses multi-element phased-array probe technology. Both technologies have been applied to the evaluation of coarse-grained components in nuclear power plants.

13.5.1   Synthetic Aperture Focusing Technique (SAFT)

SAFT for off-line data processing essentially applies the focusing effects of a large aperture probe to a data set acquired from scanning a smaller aperture probe over the area of interest [8]. In a data set acquired with a physical lens on a probe, the data are only focused in the region corresponding to the focal zone of the lens. However, SAFT processing provides a data set that is focused over nearly the entire volume of data. The technique employs a coherent summation process where the locus of points that could contribute to the center A-scan is first determined. Points along the locus are then time-shifted and summed with the center A-scan as demonstrated in Fig. 13.10 [9]. In an ideal material (homogeneous and isotropic), if a flaw signal is present, the response will be amplified because of the summation of in-phase data (coherent summation). If no flaw signal is present, the out-of-phase signals sum destructively. The process inherently produces a reduction of the background noise level. This technique has shown improvements in the evaluation of coarse-grained cast stainless steel components because of the algorithm's noise
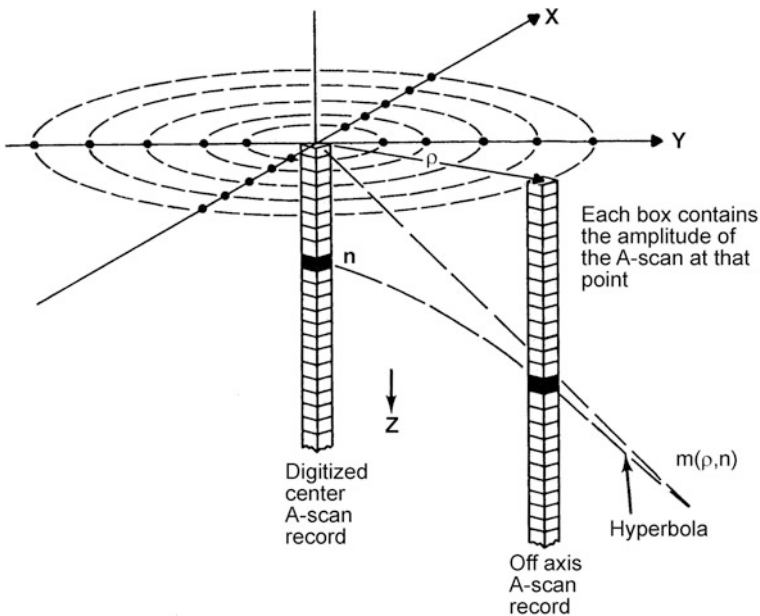
Fig. 13.10 Representation of SAFT coherent summation

reduction and improved focusing. Another application of SAFT has been in the detection of flaws on the far side of an austenitic stainless steel weld. The welding process produces large grains in the weld region and can make for a difficult inspection when passing energy through the weld to the opposite side as is necessary when geometrical constraints prohibit near-side access. In addition to enhanced flaw detection, the SAFT algorithm also provides improved flaw characterization (in terms of length and depth sizing estimates) due to its focusing effect.

An example of the benefit of SAFT processing is demonstrated in Fig. 13.11 [10], which represents data from an implanted thermal fatigue crack in the HAZ of an austenitic stainless steel piping weld. The pipe specimen was made with ASTM A-358, Grade 304 austenitic stainless steel and had an outside diameter of 610 mm (24.0 in.) and a 36 mm (1.4 in.) wall thickness. Two sections were joined with a
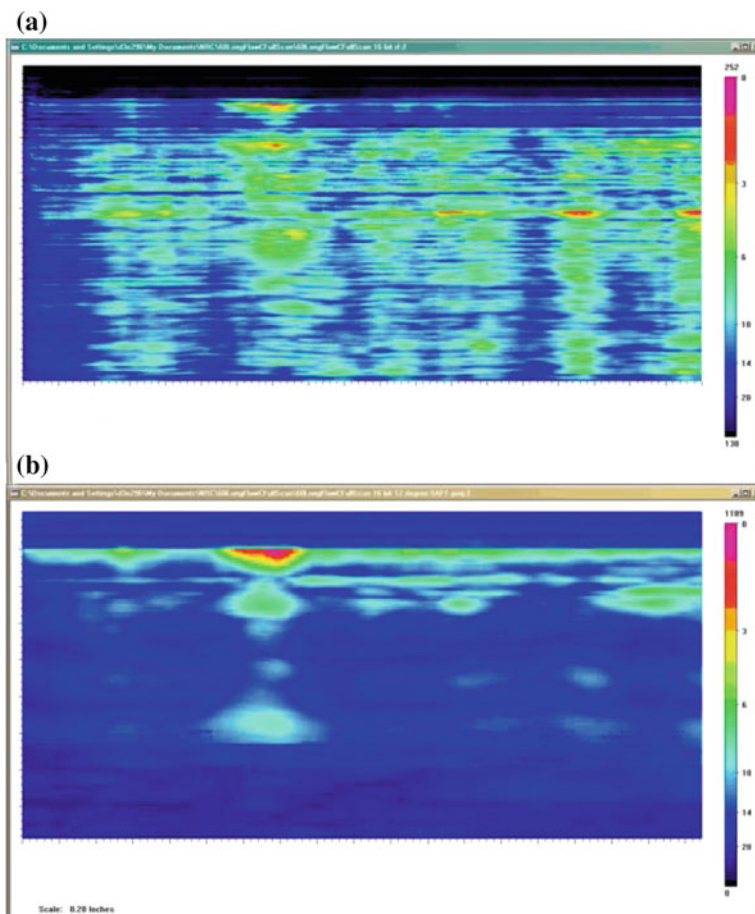
Fig. 13.11 **a** Unprocessed data from a 60 % through-wall deep thermal fatigue crack, **b** SAFT-processed image
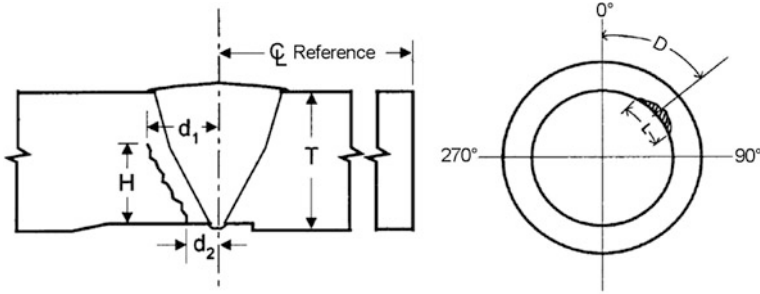
Fig. 13.12 Cross-sectional schematic in a circumferential orientation on the *left* showing weld geometry and crack location. A cross-sectional schematic in the axial orientation on the *right* shows the crack length

single-Vee girth weld using ASME SA-240, Type 304 filler material and a procedure typical of those used in nuclear applications. The component represented a primary coolant header in a boiling water reactor system. The flaw was examined from the far side of the weld, and the presence of a weld crown further limited the transducer's access to the flaw. A 60° refracted angle produced the best response from the 60 % through-wall deep crack and even so did not fully insonify the flaw. The crack length was 43.6 mm (1.72 in.). Figure 13.12 shows a schematic of the circumferentially oriented thermal fatigue crack adjacent to the weld in the HAZ. Figure 13.11a represents the acquired raw data, while Fig. 13.11b shows the SAFT-processed image. In each image, the horizontal axis represents 470 mm (18.5 in.), and the vertical axis represents 229 mm (9.0 in.) of material. The flaw is detected from these B-scan end views of the data and could be length sized. The SAFT-processed image removed much of the incoherent background noise evident in the raw data image and enhanced the flaw response so that this flaw could be detected and characterized.

13.5.2 Phased-Array Technology

Phased-array (PA) systems are composed of multi-element transducers and electronics to drive the individual elements in the probe. Each element is individually pulsed to form a beam at a defined focal region. Delays in the firing of the elements are calculated to cause the beams to add constructively and focus in the desired region. This phase shifting or phasing of the elements is the strength of PA technology in that it allows for focusing and steering of a beam. The combined delays for each individual element form a set of delay laws for a transducer and are typically computed in a software package such as Zetec's UltraVision PA Calculator. The calculator inputs include critical information on the probe design such as the number of elements, element size, and spacing. It also requires
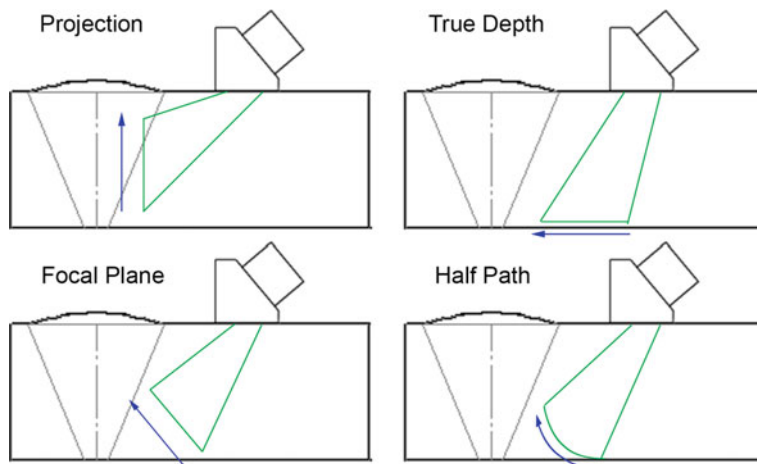
Fig. 13.13 Focus styles for azimuthal scanning

information on the wedge, including wedge angle, material, velocity, and other dimensions relative to the probe elements such as the height and offset at the middle of the first element. Information on the material being evaluated is another input to the PA Calculator and includes specimen shape, dimensions, and acoustic velocity. Additionally, the focal style and sweep of angles are specified. The typical focal styles for azimuthal beam formation (projecting in front of the probe, predominately in the primary axis) are presented in Fig. 13.13 with the green lines representing a sweep of angles, from least to greatest as shown by the blue arrows. In a projection style of focusing, an offset from a specified reference point (such as the front edge of the wedge) defines a vertical plane of focus. A true-depth focus establishes a focal plane at a constant depth from the reference point, typically the bottom of the wedge. A focal plane style of focus is established by specifying a start and stop offset and depth from the reference point. True-depth and projection focusing are special cases of focal plane focusing. Lastly, the half-path focus maintains the focal distance at a constant part path for all specified refracted angles. Half-path focusing and true-depth focusing are generally well suited for the detection of ID surface-connected flaws as observed from the typically strong corner response from such a flaw. The half-path focus additionally can provide valuable information about the upper portion of an ID-connected flaw or a mid-wall flaw and thereby allow depth sizing of a flaw. Projection and focal plane focusing can also ideally be used for mid-wall and through-wall flaw detection and sizing [11, 12].

The fundamentals of PA data analysis are discussed based on data acquired from an easily detected 4.8-mm (0.19-in.)-diameter flat-bottom hole (FBH) in a 63.5-mm (2.50-in.)-thick fine-grained stainless steel pipe section. After the basic techniques are presented on this simple flaw, two examples from data acquired in coarse-grained materials are discussed. The FBH was examined with a 1.5-MHz PA probe operating in a transmit-receive longitudinal (TRL) mode. Both the transmit
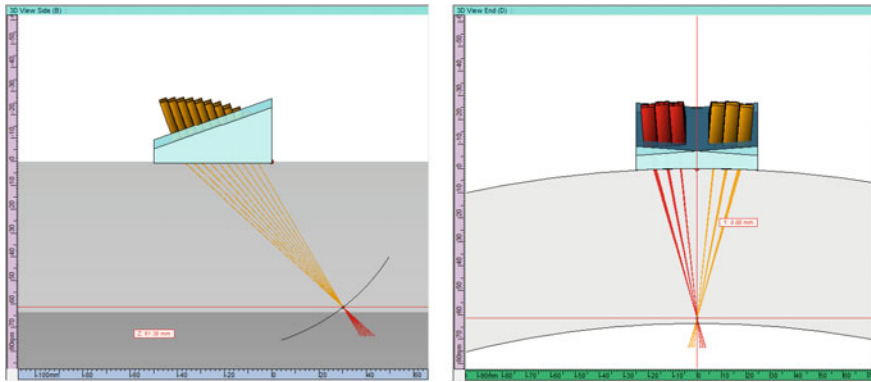
Fig. 13.14 *Side view* of the ray tracing for a half-path focus on the *left* and an *end view* on the *right*

and receive probes contained ten elements in the primary direction and three elements in the secondary direction. The PA Calculator ray trace results from a side and end view are presented in Fig. 13.14. A half-path focal style was selected with an 80 mm (3.15 in.) focal distance. The sweep of refracted angles was from 20 to 60° in 2° increments. This placed the 40° refracted angle focus at 61.3 mm (2.41 in.), close to the specimen ID as marked by the horizontal red line and black dot at the point of ray convergence in each view. Focusing at the specimen ID is recommended for the detection of ID surface-breaking flaws.

The data set acquired from raster scanning the FBH is presented in Fig. 13.15. The previously discussed views (A-scan, side, end, and top views) are identified; only the sector view is new. A sector view displays the full sweep of refracted angles, and the analyst selects the particular angle used to display the three orthogonal images—top, end, and side views. In this example, a 40° angle was chosen and is marked by the black angled line in the sector view. Additionally, the axes are color-coded with blue representing the scan axis, green representing the index axis, and magenta (and pink in the A-scan) representing time or distance into the material. The material velocity is one of the input parameters for a data file so distance or depth is easily represented in units of either length or time. Depth, as measured on the side- and end-view images, in the magenta scale, is normal to the specimen surface, whereas the A-scan depth or time is projected along the refracted angle (pink scale).

Electronic gating of the signal of interest is a valuable tool for data analysis. It allows the operator to view a slice though the data volume in an orthogonal plane and can be used to separate a flaw response from a known geometrical reflector such as a pipe counterbore in a welded region. It can also be used to isolate a flaw where multiple flaw responses are present. The gate selection and effects on image display are interconnected across the side, end, and top views as will be briefly shown in the FBH data in Fig. 13.15.
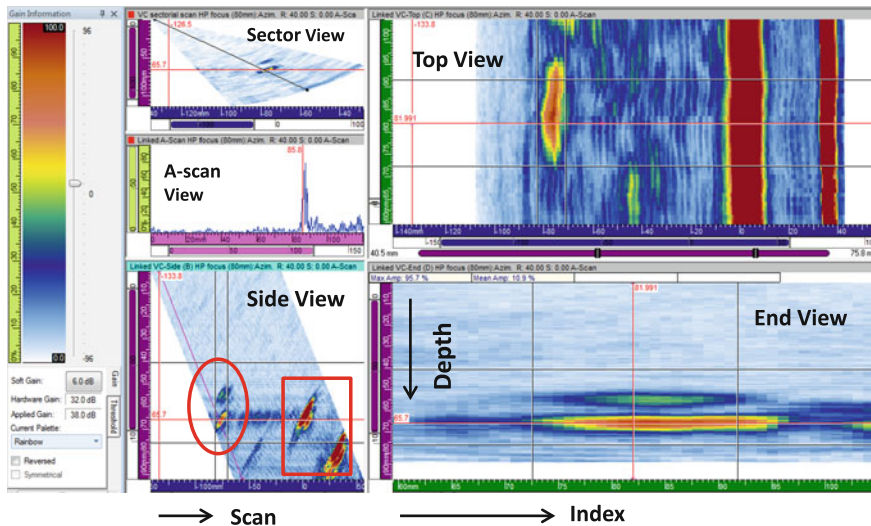
Fig. 13.15 Data images from a flat-bottom hole (*circled in red*) in a stainless steel pipe specimen

The FBH was placed 76.2 mm (3.00 in.) from the edge of the pipe. Data were acquired along a scan line that captured both the FBH response and the end-of-pipe or edge response. The side view shows the FBH response on the left (circled) and the saturated pipe edge responses on the right (boxed). Capturing the edge response in the same file as the flaw response allows one to use the edge response as a reference (or benchmark) for both position and amplitude. However, if the full volume of data is viewed from the end view, the flaw and edge responses would overlap. Electronic gating the flaw response in the side view with the black vertical lines produces an end view of that gated region only. Hence, the flaw is visible in the end view and not the edge responses. The top view represents the region gated from the end view as selected by the black horizontal lines. Because the horizontal lines capture both the flaw and the edge responses, they are all visible in the top view. For documentation purposes, the red horizontal line in the side view marks the pipe ID. A second edge response appears below the pipe ID line and represents the shear wave response. It occurs at a later time because of the slower velocity of the shear mode.

Once a flaw has been detected, it can be sized for length and for depth if a tip-diffracted or upper specular reflector response is observed. Length sizing is typically measured at a half-amplitude (−6 dB) level or at a loss of signal level (approximately −12 dB level). Depth sizing is determined from the distance between the center of the ID corner response and the center of the tip response. This process can be assisted with the display of an echo-dynamic curve along the image vertical as presented in Fig. 13.16 on the right. Additionally, the width of this machined FBH can be determined from the two responses observed from the top of
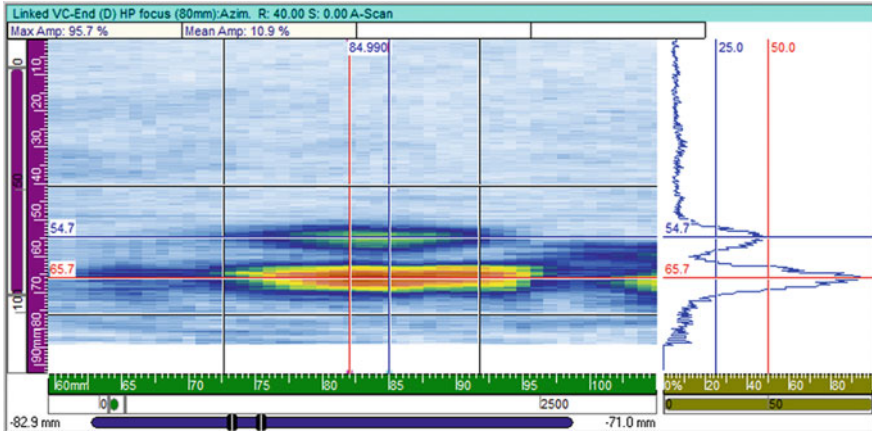
Fig. 13.16 Example of depth sizing assisted by an echo-dynamic *curve*, on the *right*

the hole in the side view. A crack with multiple branching may also exhibit more than one flaw tip with the deepest response used to report flaw depth.

Ultrasonic PA data from a crack in coarse-grained pressurizer surge line (PZR) pipe salvaged from a canceled nuclear power plant [5] are presented next. The specimen was a cast austenitic stainless steel (CASS) 304.8-mm, 12-in., Schedule 160 pipe with an approximate through-wall thickness of 33.5 mm (1.32 in.). Flaws were in situ grown in the pipe with an induction heating process that introduces and grows thermal fatigue cracks (Trueflaw Ltd.). Data were acquired with PA probes at several frequencies. A half-path focusing was selected, and focal laws generated for each probe. The left side of Fig. 13.17 depicts a ray trace of the requested sweep of refracted angles (30–60° in 1° increments) with the focal spot at the ID of the pipe at a 45° angle. The brown columns on top of the wedge represent the time delays for each of the ten elements in the primary direction (plane of the page) at the selected 45° angle. A simulated sound field density is shown on the right side of the figure at a
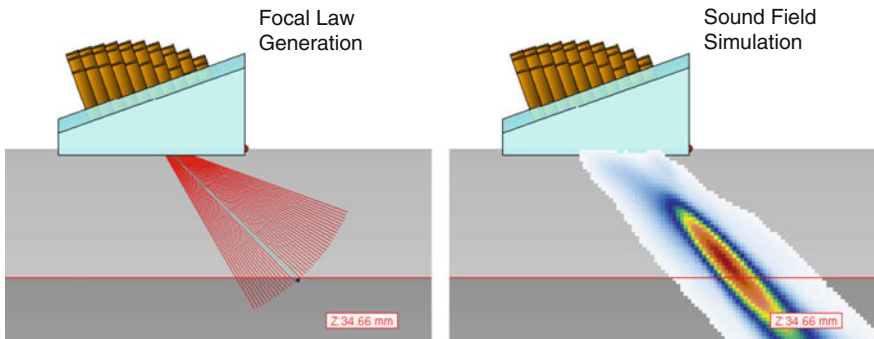
Fig. 13.17 Ray trace of specified sweep of angles, *left*, and simulated sound field, *right*
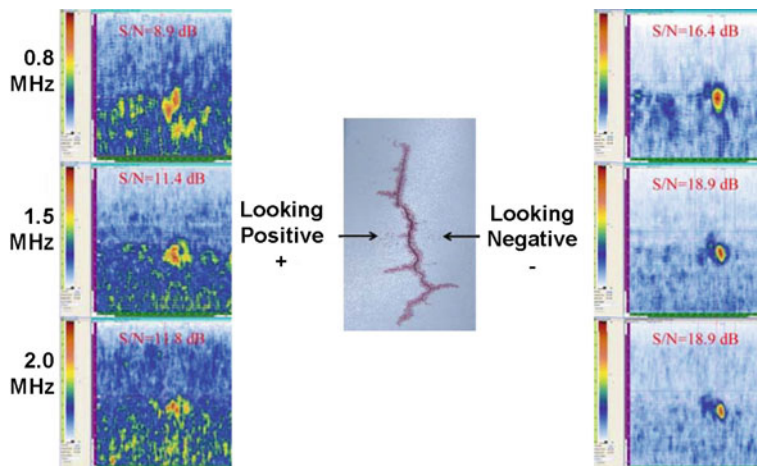
Fig. 13.18 Thermal fatigue crack responses from looking positive and looking negative sides at three probe frequencies

45° angle. Notice that the maximum simulated sound field intensity at this angle occurs above the horizontal red line, representing the ID of the pipe.

The crack was length sized at 19.3 mm (0.76 in.) based on a visual examination. Its depth was estimated at 3.4 mm (0.13 in.) or 10.1 % through wall based on validating the crack growth process in a similar material. The flaw as observed after applying a liquid penetrant red dye to enhance its visibility is shown in the middle of Fig. 13.18. The flaw was evaluated from both sides, looking positive and looking negative, with three PA probes at center frequencies of 0.8, 1.5, and 2.0 MHz. Data results are also presented in the figure, on the left for the positive view and on the right for the negative view.

Signal-to-noise (S/N) values were calculated and are displayed on the data images. The flaw has a favorable side for detection as noted by the nominal 7 dB improvement in S/N when viewed from the negative orientation. This thermal fatigue crack exhibits branching and turns, which are typical of service-induced flaws (typically intergranular stress corrosion cracking—IGSCC). Because of the flaw morphology, significantly more of the acoustic wave is reflected back to the PA probe when viewed from the negative side. A favorable response from one side of a flaw over the other is not uncommon for cracks in actual field piping; for this reason, dual-sided access to a region of interest for flaw detection is desirable. S/N values increased slightly with increasing frequency, but flaw coherency decreases, especially as viewed from the positive side.

More challenging inspections occur when tasked with flaw detection and characterization (length and depth sizing) in large-diameter (greater than 711 mm or 28 in.), thick-walled (greater than 50.8 mm or 2.0 in.) coarse-grained CASS piping as found in primary pressure boundary applications of light water reactors [13]. For these applications, even lower probe frequencies are necessary to minimize the
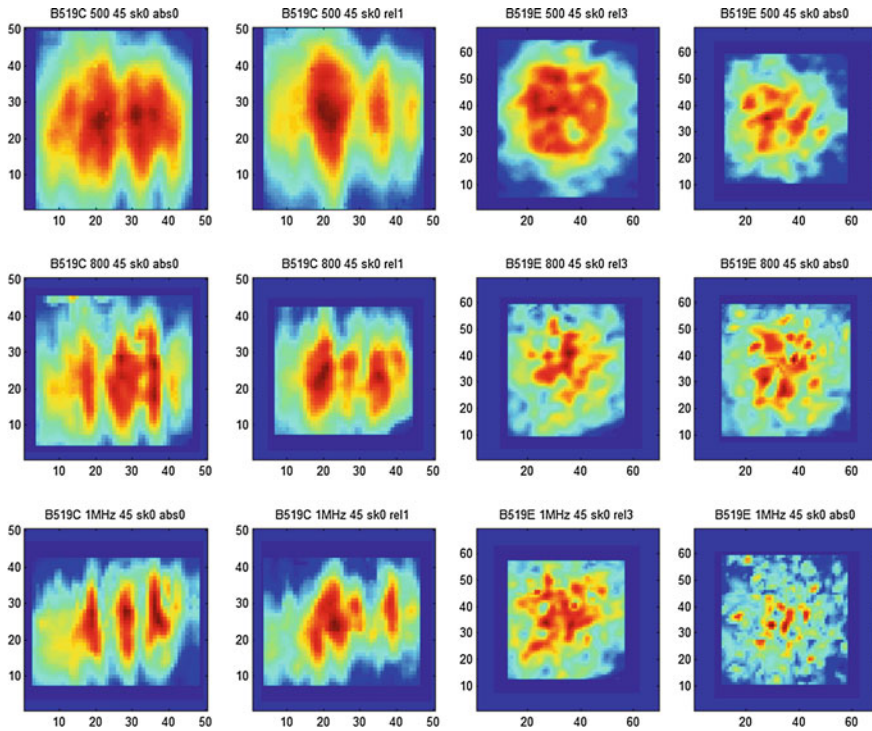
Fig. 13.19 Sound field images from columnar microstructure (*left two columns*) and equiaxed microstructure (*right two columns*). Rows represent probe center frequencies of 0.5, 0.8, and 1.0 MHz

deleterious effects of the large gains that cause scattering, beam redirection, and changes in material velocity [2]. Examples of PA sound fields formed in a columnar and equiaxed material with nominal through-wall thickness of 60 mm (2.4 in.) are displayed in Fig. 13.19. Two examples of columnar material responses are presented in the left two columns (same specimen but data points were separated by 6.4 mm or 0.25 in.) and the equiaxed responses in the right two columns (same specimen but data points were separated by approximately 19.1 mm or 0.75 in.). The top row represents the sound field form by a probe with center frequency of 0.5 MHz, the middle row represents 0.8 MHz, and the bottom row represents 1.0 MHz. The sound fields were generated with a projection style of focus coincident with the end of the specimen, while a pinducer was raster scanned over the end of the specimen to acquire the sound field image. For comparison, sound fields were generated in a fine-grained wrought stainless steel specimen of similar thickness with the results shown in Fig. 13.20 for the same three probe frequencies. These sound fields exhibit an ideal Gaussian shape with maximum sound field intensity near the center and monotonically decreasing away from center. The CASS sound field images in comparison show a lack of beam coherence.
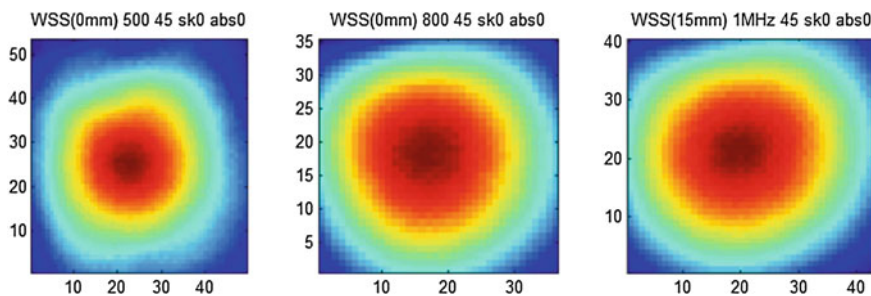
Fig. 13.20 Sound field images from fine-grained stainless steel at probe frequencies of 0.5, 0.8, and 1.0 MHz, *left* to *right*

The columnar microstructure partitioned the response with increased partition correlating with an increase in frequency, especially in the 0.5–0.8 MHz transition. Sound field scatter as measured from the images increased in the equiaxed microstructure as probe frequency increased. These data images suggest the need for a low-frequency, in the order of 0.5 MHz, evaluation of thick-walled, coarse-grained materials to reduce the effects of beam partition and scatter [3].

13.6   Summary and Future Work

Ultrasonic evaluation can provide an effective volumetric examination of welded components in nuclear power plants for the assessment of component integrity and estimates of remaining useful life. The selection of inspection parameters such as probe frequency, size, focal style in a phased-array application, and encoded scanning is necessary to ensure proper beam forming and insonification of the area of interest. SAFT processing and phased-array technologies have greatly improved the ability to detect and characterize flaws in these components. Other data acquisition methodologies including full matrix capture and post-processing of these data with a total focusing method [14] or time-reversal focusing [15, 16] are areas for future work. These techniques should continue to improve the effectiveness of an ultrasonic evaluation on a diverse range of materials.

References

1. Krautkrämer J, Krautkrämer H (1990) Ultrasonic testing of materials (4th Fully Revised Edition). Springer, New York
2. Diaz AA, Doctor SR, Hildebrand BP, Simonen FA, Schuster GJ, Andersen ES, McDonald GP, Hasse RD (1998) Evaluation of ultrasonic inspection techniques for coarse-grained materials. NUREG/CR-6594, PNNL-11171. U.S. Nuclear Regulatory Commission, Washington, D.C

3. Crawford SL, Prowant MS, Cinson AD, Larche MR, Diaz AA, Anderson MT (2014) Phased array ultrasonic sound field mapping in cast austenitic stainless steel. PNNL-23393. Pacific Northwest National Laboratory, Richland, Washington

4. Taylor TT (1984) An evaluation of manual ultrasonic inspection of cast stainless steel piping. NUREG/CR-3753, PNL-5070. U.S. Nuclear Regulatory Commission, Washington, D.C

5. Diaz AA, Cinson AD, Crawford SL, Mathews R, Moran TL, Prowant MS, Anderson MT (2011) An evaluation of ultrasonic phased array testing for cast austenitic stainless steel pressurizer surge line piping welds. NUREG/CR-7122, PNNL-19497. Pacific Northwest National Laboratory, Richland, Washington

6. Anderson MT, Diaz AA, Cinson AD, Crawford SL, Prowant MS, Doctor SR (2014) Final assessment of manual ultrasonic examinations applied to detect flaws in primary system dissimilar metal welds at north anna power station PNNL-22553. Pacific Northwest National Laboratory, Richland, Washington

7. Crawford SL, Doctor SR, Cinson AD, Watts MW, Cumblidge SE, Hall TE, Anderson MT (2011) Assessment of NDE methods on inspection of HDPE butt fusion piping joints for lack of fusion, NUREG/CR-7136, PNNL-20300. U.S. Nuclear Regulatory Commission, Washington, D.C

8. Busse LJ, Collins HD, Doctor SR (1984) Review and discussion of the development of synthetic aperture focusing technique for ultrasonic testing, SAFT UT. NUREG/CR-3625, PNL-4957. U.S. Nuclear Regulatory Commission, Washington, D.C

9. Hall TE, Reid LD, Doctor SR (1988) The SAFT-UT Real-time inspection system—operational principles and implementation. NUREG/CR-5075, PNNL-6413. U.S. Nuclear Regulatory Commission, Washington, D.C

10. Anderson MT, Diaz AA, Cinson AD, Crawford SL, Cumblidge SE, Doctor SR, Denslow KM, Ahmed S (2011) An assessment of ultrasonic techniques for far-side examinations of austenitic stainless steel piping welds. NUREG/CR-7113, PNNL-19353. U.S. Nuclear Regulatory Commission, Washington, D.C

11. Drinkwater BW, Wilcox PD (2006) Ultrasonic arrays for non-destructive evaluation: a review. NDT&E Int 39:525–541

12. Olympus NDT (2004) Introduction to phased array ultrasonic technology applications: R/D tech guideline. R/D tech part number DUMG068C. Olympus NDT, Waltham, Massachusetts

13. Ruud CO, Diaz AA, Anderson MT (2009) Grain structure identification and casting parameters of austenitic stainless steel (CASS) Piping, PNNL-19002. Pacific Northwest National Laboratory, Richland, Washington

14. Zhang J, Drinkwater BW, Wilcox PD (2011) Effects of array transducer inconsistencies on total focusing method imaging performance. NDT&E Int 44:361–368

15. Edelmann GF, Lingevitch JF, Gaumond CF, Fromm DM, Calvo DC (2007) Comparison of a Subrank to a full-rank time-reversal operator in a dynamic ocean. J Acoust Soc Am 122 (5):2706–2714. doi:10.1121/1.2783127

16. Gliozzi AS, Griffa M, Scalerandi M (2006) Efficiency of time-reversed acoustics for nonlinear damage detection in solids. J Acoust Soc Am 120(5):2506–2517

Chapter 14
Nondestructive Visualization Using Electromagnetic Waves for Real and Practical Sensing Technology for Robotics

Hiroyoshi Togo, Soichi Oka, Yoshihisa Fujii and Yuko Fujiwara

Abstract This section describes novel methods of electromagnetic wave nondestructive visualization (NDV) for assessing qualification and durability of concrete and wooden structures, which involves devices, systems, and image processing. As the first, the basic knowledge and principles on dielectric properties of materials, wave propagation in media involving plane waves in vacuum and in non-conducting and non-magnetic dielectric media were introduced. As the second, the dielectric properties of concrete and NDV techniques for concrete structures were introduced. After the introduction of conventional methods to detect internal cracks in concrete structures, a novel development of a millimeter wave scanner for NDV of concrete was introduced where the performance of the scanner to detect surface cracks covered by other sheet materials was discussed. Miscellaneous image processing techniques to recognize the target using pattern recognition methods were also introduced. Wood is a material made of the plant cells of trees. Wood shows anisotropy in physical and mechanical properties, such as elastic moduli, strength, and dielectric constants. In addition, wood is deteriorated by biological agents such as insects and fungi, and these deterioration often generate in inner or hidden areas of wood and wooden construction. The deterioration is closely associated with moisture content of wood. In this section, the feasibility of technologies using electromagnetic waves for the nondestructive evaluation of properties and deterioration of wood and wooden construction, is introduced. The reaction of wood to electromagnetic wave such as transmission and reflection of millimeter wave through and from wood was discussed. The development of wooden wall scanner for nondestructive diagnoses of wooden houses using FMCW radar was introduced.

H. Togo (✉) · S. Oka
NTT Microsystem Integration Laboratories, 3-1, Morinosato Wakamiya,
Atsugi 243-0198, Kanagawa, Japan
e-mail: togo.hiroyoshi@lab.ntt.co.jp

Y. Fujii · Y. Fujiwara
Laboratory of Wood Processing, Graduate School of Agriculture, Kyoto University,
Oiwake-cho Kitashirakawa, Sakyo, 606-8502 Kyoto, Japan

Contents

14.1 General Introduction

Imaging techniques for investigating physically inaccessible objects have been topics of research for many years and have found widespread applications in the field of nondestructive visualization (NDV). All electromagnetic wave methods in NDV involve Maxwell's equations and cover a broad range of the electromagnetic spectrum, from static or direct current to high frequencies, e.g., X-ray and gamma ray methods. Imaging science is concerned with the formation, collection, duplication, analysis, modification, and visualization of images. Electromagnetic NDV is essential for detecting anomalies or defects in both conducting and dielectric materials by generating two-dimensional or three-dimensional image data based on the electromagnetic wave principles. For forward imaging approaches, the excitation transducers usually couple the electromagnetic energy to the test objects, while the receiving sensors measure the response of energy/material interaction. Depending on different energy types and/or levels, various electromagnetic sensors/transducers can be used for a broad range of applications, e.g., microwave imaging and millimeter wave imaging. After acquiring and storing the electromagnetic images, those data are passed through the inversion-techniques block, which involves object reconstruction, pattern recognition, etc. Electromagnetic wave NDV has been attracting much attention in applications for the evaluation of concrete and wood structures, e.g., buildings, bridges, and houses. Concrete and wood consist of many dielectric components, which provide complexly electromagnetic responses of transparency, reflection, refraction, and scattering. This section describes novel methods of electromagnetic wave NDV for assessing

qualification and durability of concrete and wood structures, which involves devices, systems, and image processing.

14.2 Basic Principles

The foundations of nondestructive visualization (NDV) using electromagnetic (EM) waves lie in EM theory, and the history of this field spans more than two centuries, which is based on the subject of numerous texts such as those by Jackson [1] and Smythe [2]. This overview outlines the basic building blocks, especially concrete, needed to work quantitatively with NDV using EM waves.

Maxell's equations mathematically describe the physics of EM fields, while constitutive relationships quantify material properties. Combining the two provides the foundations for quantitatively describing EM wave signals.

14.2.1 Dielectric Properties of Materials

A dielectric material conducts EM waves and is commonly referred to as an insulator. The transition from insulator, which is characterized by low electric conductivity, to conductor, which has high conductivity, is gradual without any clear cutoff point, and many existing materials exhibit both properties. If the conductivity of a material is high, it attenuates EM waves to a larger extent, resulting in a lower penetration depth. The two significant properties of a dielectric material are the real and imaginary parts of its complex dielectric permittivity. The real part is associated with the phase velocity, and the imaginary part signifies the conductivity or attenuation of EM waves in the medium. Dielectric permittivity is often referred to as the dielectric "constant." This term is misleading, since this property of a dielectric material varies due to several factors. Bell et al. [3] provided some insight into these variations through underlying molecular and atomic mechanisms associated with dielectric constant and conductivity of such materials. A dielectric material increases the storage capacity of a capacitor by neutralizing charges at the electrode surfaces. This neutralization can be imagined to be the result of the orientation or creation of dipoles opposing the applied field. Such a polarization is proportional to the polarizability of the dielectric material, that is, the ease with which it can be polarized. Polarizability is defined as the average induced polarization per unit field strength. The greater the polarizability of a material, the greater will be its dielectric constant. There are four recognized mechanisms of polarization: (a) electronic, (b) atomic, (c) orientation, and (d) space charge or interfacial. The first three polarization mechanisms are forms of the dipole polarization previously mentioned.

They result from the presence of permanent dipoles or from dipoles induced by external fields. The fourth type—space charge polarization—results from charge

carriers that are usually present and more or less free to move through the dielectric. When such carriers are impeded in their movement and become trapped in the material or at interfaces, space charge concentrations result in a field distortion that increases the capacitance. The above discussions indicate that the dielectric constant of a material is not such a constant, but is a function of polarizability, which is in turn a function of frequency, temperature, local fields, applied field strength, availability and freedom of charge carriers within the dielectric, and local field distortions. Dielectric conductivity is also affected by the same factors. This is because dielectric conductivity is a function not only of ohmic conductivity but also of the power consumed in polarizing the material. In addition to the "intrinsic loss" caused by the conduction process, there could also be "scattering loss" due to the presence of inhomogeneities within the medium. Since the aggregate particle size is much smaller than the EM wavelength considered at a microwave range, the inhomogeneities in concrete are not likely to cause significant scattering losses. However, this is not true for inhomogeneities with large dimension (e.g., delaminations) and large conductivity (e.g., rebar) or with application at the millimeter wave (MMW) range. These inhomogeneities will be modeled separately, and their effect will be taken into account for the purpose of waveform synthesis.

14.2.2 Wave Propagation in Media

One of the most important consequences of Maxwell's equations is the equations for EM wave propagation in a linear medium. In the absence of free charge and current densities, Maxwell's equations are

$$
\begin{aligned}
\nabla \cdot D &= 0 \\
\nabla \times H &= \frac{\partial D}{\partial t} \\
\nabla \cdot B &= 0 \\
\nabla \times E &= \frac{\partial B}{\partial t}
\end{aligned}
\tag{14.1}
$$

The wave equations are derived by taking the curl of

$$
\begin{aligned}
\nabla \times \nabla \times E &= -\nabla \times \frac{\partial B}{\partial t} \\
\nabla \times \nabla \times H &= -\nabla \times \frac{\partial D}{\partial t}
\end{aligned}
\tag{14.2}
$$

For uniform isotropic linear media, we have $D = \varepsilon E$ and $B = \mu H$, where ε and μ are in general complex functions of frequency ω. Then, we obtain

$$\nabla \times \nabla \times E = -\varepsilon\mu\frac{\partial^2 E}{\partial t^2}$$

$$\nabla \times \nabla \times B = -\varepsilon\mu\frac{\partial^2 B}{\partial t^2}$$

(14.3)

Since $\nabla \times \nabla \times E = \nabla(\nabla \cdot E) - \nabla^2 E = -\nabla^2 E$ and,   similarly, $\nabla \times \nabla \times B = -\nabla^2 B$,

$$\nabla^2 E = \varepsilon\mu\frac{\partial^2 E}{\partial t^2}$$

$$\nabla^2 B = \varepsilon\mu\frac{\partial^2 B}{\partial t^2}$$

(14.4)

Monochromatic waves may be described as waves that are characterized by a single frequency. Assuming the fields with harmonic time dependence $e^{-i\omega t}$, so that $E(x,t) = E(x)e^{-i\omega t}$ and $B(x,t) = B(x)e^{-i\omega t}$, we obtain the Helmholtz wave equations

$$\nabla^2 E + \varepsilon\mu\omega^2 E = 0$$

$$\nabla^2 B + \varepsilon\mu\omega^2 B = 0$$

(14.5)

Plane waves in vacuum
Suppose the medium is vacuum, so that $\varepsilon = \varepsilon_0$ and $\mu = \mu_0$. Furthermore, suppose $E(x)$ varies in only one dimension. Then, Eq. 14.5 becomes

$$\frac{d^2 E(z)}{dz^2} + k^2 E(z) = 0$$

(14.6)

where the wave number $k = \omega/c$. This equation is mathematically the same as the harmonic oscillator equation and has solutions. Therefore, the full solution is

$$E_k(z,t) = \alpha e^{\pm ikz - i\omega t} = \alpha e^{-i\omega(t\mp\frac{z}{c})}$$

(14.7)

where α is a constant vector.

This represents a sinusoidal wave traveling to the right or left in the z-direction at the speed of light. Using the Fourier superposition theorem, we can construct a general solution of the form

$$E(z,t) = A(z-ct) + B(z+ct)$$

(14.8)

Plane waves in non-conducting, non-magnetic dielectric
In a non-magnetic dielectric, we have $\mu = \mu_0$ and the index of refraction

$$n(\omega) = \sqrt{\frac{\varepsilon(\omega)}{\varepsilon_0}} \qquad (14.9)$$

The results are the same as in vacuum, except that the velocity of wave propagation or the phase velocity is now $v = c/n$ instead of c. Then, the wave number is

$$k(\omega) = n(\omega)\frac{\omega}{c} \qquad (14.10)$$

14.3 NDV for Diagnosis of Concrete Structures

This section reviews some NDVs for diagnosis of concrete structures and introduces MMW imaging technology for detecting submillimeter-wide cracks under a bill-posting prevention sheet on a concrete telegraph pole.s

14.3.1 Degradation Phenomenon of Concrete Structures

In the social environment based on concrete structures, the maintenance and renewal of aging infrastructure is becoming a major issue. Operating such aged facilities efficiently and safely requires the following maintenance cycle: design, inspection, diagnosis, evaluation, and implementation of countermeasures. To assure the safety of facilities, it is necessary to ascertain their state to a high degree of accuracy and detail. Moreover, among the steps in the maintenance cycle, the inspection and diagnosis processes are extremely important. At present, the usual way to inspect and diagnose facilities with high precision is to perform strength testing and material analysis on samples removed from an existing facility. This method, by definition, consumes part of an existing facility, so it is difficult to apply it to the entire structure of an aged existing facility. Furthermore, in large-scale facilities like tunnels, the conditions often vary with the measurement position, so it is difficult to evaluate the condition of the whole facility from local measurements. On the other hand, non destructive testing evaluates the condition of a structure without damaging it, so it has two advantages: it can be applied to the whole of a large-scale structure and can efficiently evaluate the degradation condition of that structure. The ideal situation and the actual state of the maintenance of steel-reinforced concrete structures are shown schematically in Fig. 14.1 [4] using the analogy between the degradation process of steel-reinforced concrete and the progress of a disease in humans. In the case of a human disease, if the disease can be detected in its latent phase (i.e., its premorbid phase) by regular health checks, it can be prevented from reaching an acute state, so major surgery is unnecessary. Likewise, in the case of concrete structures, it is important to take effective action at
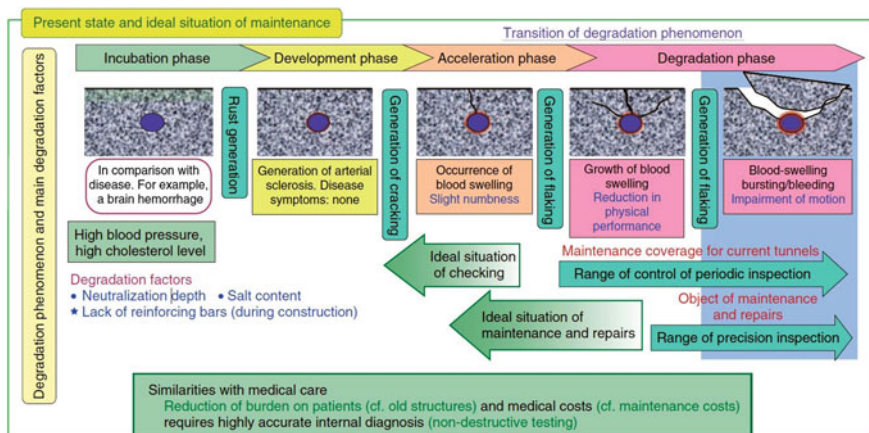
Fig. 14.1 The ideal situation and the actual state of the maintenance of steel-reinforced concrete structures

the stage before visible degradation of the concrete occurs. If high precision facility evaluation by non-destructive testing were available, it would be possible to continue operating and maintaining existing facilities for a long time.

14.3.2 NDV Techniques for Concrete Structures

Methods for seeing through the covering material to observe the condition of the surface beneath it are classified by the physical phenomena that are used. In particular, methods that use X-rays, ultrasonic waves, or microwaves have been applied in various fields. Table 14.1 lists the series of NDV techniques for each target of concrete structures, as shown in Fig. 14.2.

X-rays are a form of ionizing radiation that have high transmittance for materials other than water and heavy metals and allow imaging with micrometer-order spatial resolution. They are applied for medical purposes such as 3D imaging of internal organs and bones, but since they are harmful to humans, exposure levels must be carefully controlled: This would be a safety problem for workers using X-rays inspecting multiple concrete poles every day. Another problem is that practical X-ray imaging equipment is based on transmission, so it should really have a separate transmitter and receiver; however, for work sites with limited work space, reflection-based imaging equipment can be used.

Ultrasonic waves are elastic waves that have almost no effect on the human body and have high transmittance and millimeter-order spatial resolution for reflection-based imaging. They are therefore used for medical three-dimensional (3D) imaging of fetuses, but the procedure requires the application of grease between the ultrasonic wave transceiver and object to be observed for efficient

Table 14.1 Evaluation methods and NDT for reinforced concrete structures

Inspection items in concrete structure		NDT
Structure dimensions	Position of reinforcing bar	EM induction, EM wave, ultrasonic wave, or X-ray
	Thickness of covering concrete on reinforcing bar	EM induction, EM wave, or ultrasonic wave
	Diameter of reinforcing bar	EM induction, EM wave
	Thickness of concrete	Elastic or ultrasonic wave
Deterioration situation	Delamination and internal cavity	Hammering, infrared camera, elastic wave, EM wave, ultrasonic wave, or X-ray
	Surface crack	Image processing, laser scanning, elastic wave, EM wave, ultrasonic wave
	Compression of strength	Hammering or ultrasonic wave
	Correction of reinforcing bar	Half-cell potential, polarization resistance, or ultrasonic wave
	Carbonation and salt-induced deterioration	EM wave

Fig. 14.2 Evaluation methods and NDT for reinforced concrete structures. a. Structure dimension b. Deterioration situation
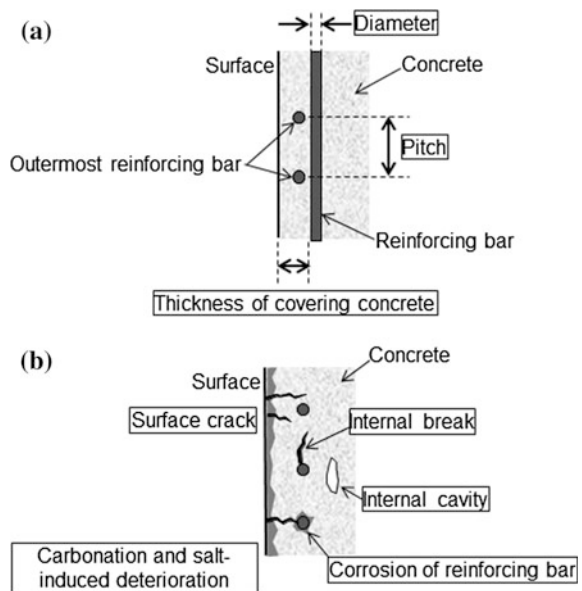
sound wave transmission. The expense of applying and removing grease would be a problem for inspection of a huge number of poles.

Microwaves are EM waves that have high transmittance and centimeter-order spatial resolution inside concrete, so they are used to inspect rebar. However, this technique does not provide the submillimeter-order spatial resolution required for crack detection, so imaging performance remains a problem.

14.3.2.1 Conventional Electromagnetic Wave Techniques for Internal Checking

The continued deterioration of concrete structures, such as buildings, bridges, tunnels, and highways, increases the danger of their collapse. The durability of concrete structures can be estimated from the position, number, and diameter of their reinforcing bars. The distance from the surface of the structure to a reinforcing bar is a particularly important parameter for estimating the bar's corrosion state. Pulse radar, which makes a two-dimensional (2D) azimuth–range image, has attracted much attention for nondestructively checking the positions of reinforcing bars because of its simple operation [5]. However, a conventional pulse radar system cannot measure the positions of shallow bars due to its low depth resolution of several tens of millimeters defined from its nanosecond-order pulse width. The measurement of the shallow position is becoming important because the shallow bar is easily corroded.

An ultra-wideband (UWB) antenna is an attractive solution to improve the depth resolution because of its wide bandwidth corresponding to the pulse width of a few hundred ps [6–10]. However, conventional UWB antennas are not suited for measuring the bar's position in reinforced concrete structures because of its omnidirectional radiation which induces multipath reflected pulse waves as a noise signal. This has led to the development of tapered slot antennas (TSAs), which are directional UWB antennas with a directivity of over 6 dBi [11–20]. The TSAs usually have isotropic radiation characteristics in the E- and H-planes, which evenly illuminate the target. However, the isotropic radiation characteristics are inefficient regarding the illumination power for a synthetic aperture radar (SAR) technique which is a post signal processing technique used to improve the resolution of the radar image. This technique uses the correlation between the acquired 2D data along the azimuth axis. The spatial resolution of the SAR image improves the signal-to-noise ratio (SNR), resulting in a higher contrast image. The SAR technique requires a wide illuminated area in the H-plane and a narrow one in the E-plane of the antenna because the correlation is performed only for the H-plane data. Therefore, we developed a TSA with orthogonally anisotropic directivity, that is, wide and narrow radiation patterns in the H- and E-planes, respectively.

The aim of our study was to develop an SAR system with high spatial resolution for the nondestructive evaluation of reinforced concrete structures. We numerically and experimentally investigated a new antenna suitable for a SAR system using an ultra-short pulse. Testing of a prototype system demonstrated that it can detect the bars in reinforced concrete samples.

Mochizuki optimized the shape of a Vivaldi tapered slot antenna (VTSA) by using a commercially available numerical simulator [21]. Using this simulator makes it easy to evaluate EM fields (both near and far) from an antenna and its S-parameters.

Figure 14.3 illustrates the configuration of our developed antenna, which is an anisotropic directional antenna (a VTSA with substrate cutting and corrugation). It is fabricated on a substrate with a relative dielectric constant of 3.75, a loss tangent
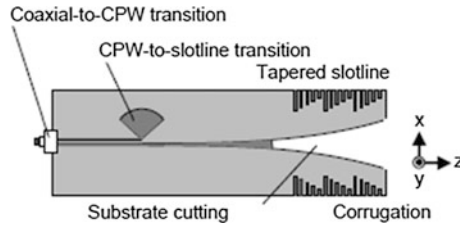
Fig. 14.3 Anisotropic directional antenna
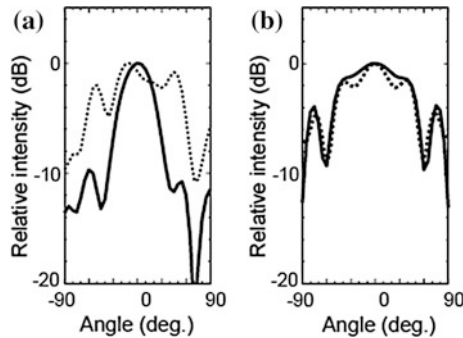
Fig. 14.4 Radiation patterns of anisotropic directional antenna (*solid line*) and VTSA without substrate cutting and corrugation (*dotted line*) in E-plane (**a**) and H-plane (**b**)

of 0.02, and a thickness of 1.0 mm including the thin copper coating. It consists of three parts: a coaxial-to-coplanar waveguide (CPW) transition, a CPW-to-slotline transition, and a tapered slotline. It has a uniquely shaped substrate below the slotline, which is formed by substrate cutting. This shape produces the difference in radiation patterns in the E- and H-planes (which are defined by the xz and yz planes, respectively). Corrugation is used on both sides of the tapered slotline to reduce sidelobe levels. Mochizuki computed the radiation patterns in the E- and H-planes of the developed antenna as well as those of an isotropic directional antenna (VTSA without substrate cutting and corrugation). Figure 14.4 shows their radiation patterns at 10 GHz. The developed antenna has a narrower beam width for the E-plane and a wider beam width for the H-plane than the isotropic directional antenna. The ratio of the 10-dB beam width of the H-plane to that of the E-plane for the developed antenna is 120:60 (2:1), which is higher than that for other TSAs. This difference in the radiation patterns is mainly due to removing the substrate below the tapered slotline. The EM energy flowing from the feed point to the aperture of the TSA is highly concentrated in the substrate below the tapered slotline. When the substrate below the slotline is removed, the energy spreads in the direction of the H-plane; it cannot spread in that of the E-plane due to the limited slotline width. Calculation of the antenna gain shows that the corrugation and substrate cutting increased the gain from 4.7 to 7.6 dBi.
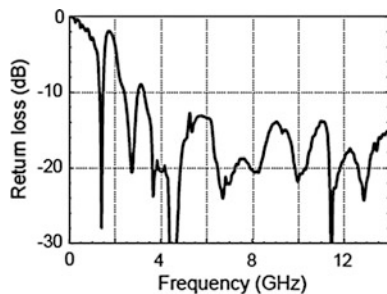
Fig. 14.5 Return loss (S11) of anisotropic directional antenna
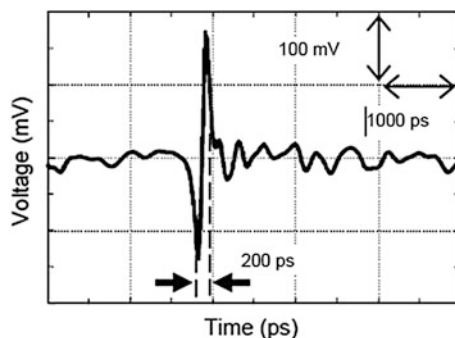
Fig. 14.6 Time-domain waveform transmitted with anisotropic directional antennas in free space at the distance between antennas was 150 mm

The return loss (S11) of the developed antenna, shown in Fig. 14.3, was measured with a vector network analyzer in free space, and the results are shown in Fig. 14.5. It was found that this antenna works well at a higher frequency than 3.2 GHz.

A pulse transmission test in free space (distance of 150 mm) between two of the developed antennas was conducted. The transmitting antenna was driven with a Gaussian pulse with a pulse width of 100 ps. The waveform of the received pulse in the time domain was measured with a high-speed digital sampling oscilloscope. The results, shown in Fig. 14.6, reveal a monocycle pulse wave with a width of 200 ps, which corresponds to a bandwidth of over an octave or more.

The SAR imaging system we developed is based on a bistatic pulse radar and uses two of the developed anisotropic directional antennas, as shown in Fig. 14.7. Sinusoidal input for the comb generator is supplied from a signal generator through the amplifier. A pulse with a width of 100 ps is generated and radiated through the transmitting antenna to the target. The backscattered pulse from the target is obtained in the time domain with a high-speed digital sampling oscilloscope through the receiving antenna.

Fig. 14.7 Pulse radar system
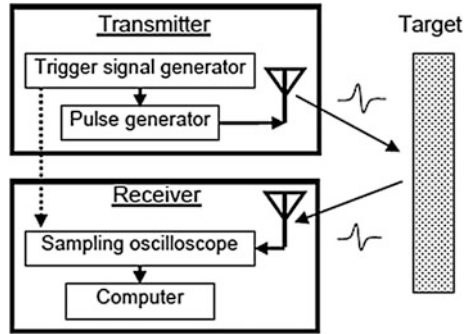
Fig. 14.8 Experimental setup for SAR imaging
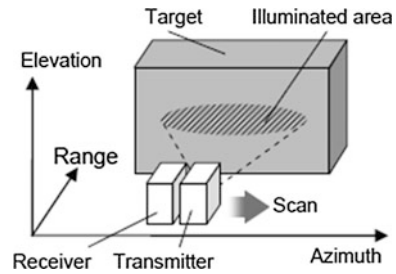
Fig. 14.9 Radar image of reinforced concrete sample
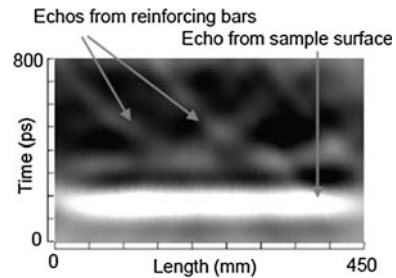
The setup used for testing the SAR system is shown in Fig. 14.8. The azimuth and range axes were parallel to the scan and pulse radiation directions, respectively. The elevation axis was normal to the azimuth and range axes. The H- and E-planes of the antennas corresponded to the azimuth range and elevation range planes, respectively, so the illuminated area was elliptical with the major axis parallel to the azimuth axis, as shown in Fig. 14.9.

The testing was done using a reinforced concrete sample with two reinforcing bars. As shown in Fig. 14.10, one bar had a radius of 10 mm and the other had a radius of 12 mm. They were set 13 and 11 mm deep, respectively, too shallow to be detected with a conventional imaging system because of its poor resolution (more than 25 mm).

Fig. 14.10 SAR image of reinforced concrete sample
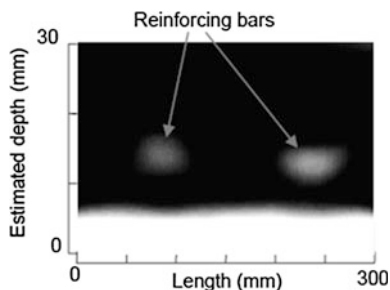
The transmitter illuminated the sample with a pulse wave, and the receiver recorded the backscattered pulse. This process was performed repeatedly and continuously, while the system was moved over the sample. After the sample had been scanned, a 2D radar image corresponding to the azimuth range plane was displayed on a personal computer (PC).

An SAR image, which is an image with higher spatial resolution than the radar image, was also obtained by performing SAR signal processing on the personal computer. This SAR system uses a conventional SAR imaging technique based on the delay-and-sum approach, which collects the backscattered data with a curved profile at a point in the radar image. The number of collected data points for a curved profile was 400, and the relative permittivity of the sample was assumed to be 9.

14.3.3   Millimeter Wave Scanner for NDV

14.3.3.1   Features in Millimeter Wave Technologies

To overcome the issues of safety, economy, and imaging performance, we took up the difficult R&D challenge of using MMWs, which are EM waves of higher frequencies than micrometer waves, to develop improved imaging technology. The MMW band includes frequencies from 30 to 300 GHz, but the frequencies specified by the radio law for radar use (imaging is also one use of radar) range only from 76 to 77 GHz (the band allocated to MMW radar). The allocation of MMW radar band frequencies for inspection purposes is being studied. The main concerns in developing MMW radar technology for detecting cracks in CP under bill-posting prevention sheets are (1) the transmittance of the plastic sheet, (2) submillimeter spatial resolution, and (3) ease of equipment operation.

In the band allocated for MMW radar, the relative permittivity of polyvinyl chloride is about 3 and the dielectric tangent is about 0.01, so if Oka and Togo assume a planar wave incident on a 4-mm-thick infinite plane, the attenuation due to reflection and propagation loss is low, about −10 dB, so transmittance is not a significant problem for equipment design. There is concern, however, that the texture of the bill-posting prevention sheet may have a greater effect on attenuation than the sheet's material properties because of wave scattering at the surface.
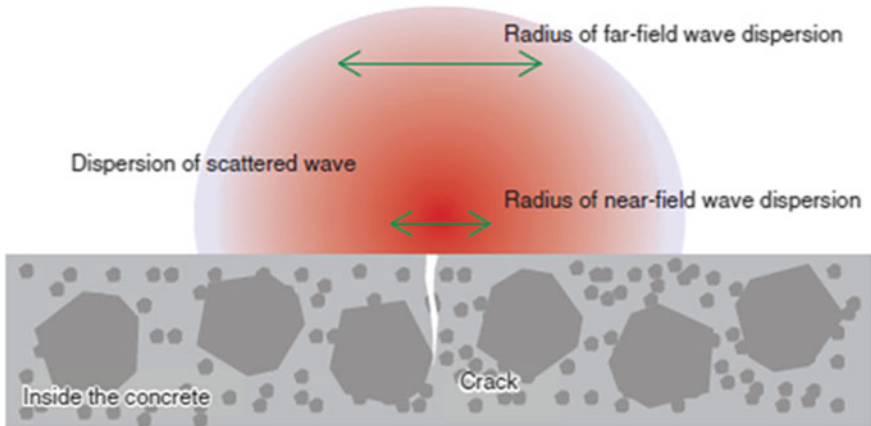
Fig. 14.11 Dispersion of wave scattering from a crack

Another factor is that the width of the cracks to be detected is less than 1/10 the wavelength of the band allocated to MMW radar, so considering the radar cross-sectional area, the backscattering will be less than −30 dB relative to the power irradiating the surface. On the basis of the above estimates, we are investigating an equipment design in which the irradiation power of the MMW radiation is set to −50 dB to pass through the bill-posting prevention sheet to reach the crack and then pass back through the sheet as backscattering.

Because the minimum beam diameter obtainable by focusing with a quasi-optical system using a lens is limited by diffraction, the beam width is limited to approximately the wavelength. The spatial resolution is therefore about half the wavelength, which means that it is difficult to detect submillimeter-wide cracks using the approximately 4-mm wavelength of the band allocated to MMW radar. For this reason, we have been trying to improve the spatial resolution by detecting near-field scattering. The concept, as shown in Fig. 14.11, is that spatial resolution can be improved because the waves scattered from the crack, which can be regarded as a minute point in a cross section that includes the crack width direction and crack depth direction, are spherical waves. The dispersion of the scattered waves is thus small in the near field. Detecting near-field scattering with submillimeter-order spatial resolution requires the use of an antenna whose aperture is equivalent to the width of the scattered waves at the detection height. Oka and Togo are therefore investigating aperture design at the antenna placement height with consideration given to operability.

Inspecting all the concrete telegraph poles in the entire country for structural integrity requires highly efficient work. The shortest imaging time could be achieved by covering the entire bill-posting prevention sheet with the antenna for detecting near-field scattering that is needed in the detection of submillimeter-wide cracks, but it is difficult to arrange antenna elements at submillimeter intervals in the crack width direction. Even if the elements are arranged at millimeter intervals,

since the bill-posting prevention sheet is about 1.5 m long in the vertical direction, the number of antenna and MMW transmitter and receiver components would be huge and the equipment cost would be uneconomical. The equipment would also be large, which would reduce its portability. To maintain economy and portability, we considered a method in which ten antennas are arranged in a one-dimensional (1D) array in the crack length direction and the array is scanned in the crack width direction. Spatial resolution in the crack width direction is maintained by detecting scattered waves at submillimeter intervals. The equipment can perform multiple scans in the crack length direction to image the entire surface.

14.3.3.2   Application of Millimeter Wave Scanner

Nippon Telegraph and Telephone Corporation (NTT) in Japan expends much effort in the long-term maintenance and management of the communication network to provide safe and secure communication services to customers throughout the country. Concrete telegraph poles, which are perhaps the most familiar of those facilities to most of us, are susceptible to physical damage, so a high degree of structural integrity is required. Concrete poles and other structures degrade over time; therefore, integrity maintenance requires periodic inspections so that degraded structures can be repaired. Since NTT owns more than ten million CP, such inspection is a daily job for a large number of employees.

Concrete poles carry cables, transformers, and other equipment, which produce an unbalanced load on the structure over many years. The resulting bending stress produces fatigue cracking on the pole surface. When such cracking penetrates as far as the rebar inside the pole, rainwater may enter and cause corrosion that can lead to the rebar breaking. Steel rebar, which is strong in tension, is vital to balance the concrete, which is strong in compression but weak in tension, so broken rebar reduces the integrity of a CP. Considering this degradation mechanism, we focused on the initial stage of degradation, which is the appearance of cracks on the concrete surface. NTT has experimentally verified that submillimeter-wide surface cracks have a high probability of reaching the rebar, so early discovery of the initial degradation is considered to be the best approach to structural integrity maintenance. Therefore, careful visual inspection of the pole surface is performed.

The NTT Access System Laboratories is developing a new technique to replace human visual inspection by using software to automatically detect cracks from images acquired with a digital camera. More details of this technique are given in the Feature Article "Detection of Cracks in Concrete Structures from Digital Camera Images," in this issue [22].

To prevent unauthorized posting of advertisements on CPs, 2-mm-thick sheets of textured polyvinyl chloride (bill-posting prevention sheets) are bonded to the poles with acrylic resin in the region within hand reach (Fig. 14.12). Exposing the covered surface for visual inspection would require significant work to strip off the sheet and then replace it with a new one after inspection. Since that would not be economically feasible, Oka and Togo conducted research and development (R&D)

Fig. 14.12 Bill-posting
prevention sheet on a concrete
pole

of a nondestructive inspection technique that allows observation of the pole surface
condition without the need to remove the bill-posting prevention sheet [23].

On the basis of the design policy described above, we constructed a portable
device called CP scan for detecting cracks in CPs under bill-posting prevention
sheets (Fig. 14.13). Sixteen intensity phase detection modules having tapered slot
antennas are arranged in a 1D array. A signal from the module is input to the
baseband circuit at submillimeter scanning intervals in sync with a signal from an
encoder attached to the wheels of the unit: In this way, the condition of the CP
surface is imaged. The case containing the module, encoder, and baseband circuit
weighs less than 4.1 kg and can be held in one hand for scanning. The entire
system, including the power supply and PC, weighs less than 10 kg and can be
carried by a single person. The 16 modules can image an 8-cm-wide path in a single
scan, and one or two workers can inspect a pole in about three minutes by making
five or six scans to image the entire region of the bill-posting prevention sheet
(Fig. 14.14).

Fig. 14.13 CP scan equipment
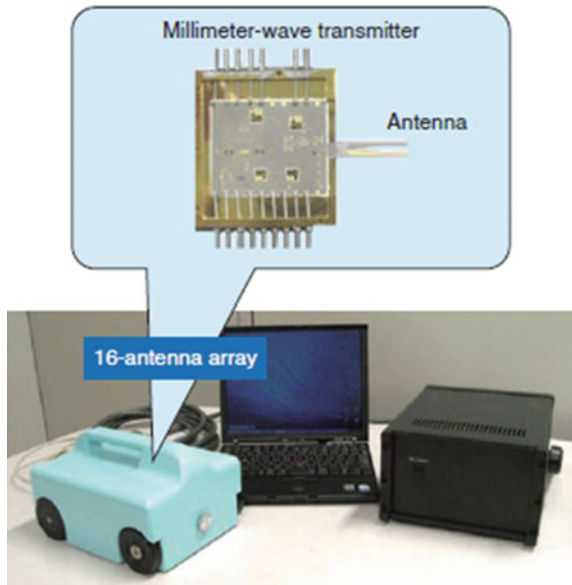
Fig. 14.14 Schematic illustration of CP scan in use

14.3.3.3   Surface Crack Detection with Millimeter Wave Scanner, Crack Scan

To test the practicality of CP scan, a bill-posting prevention sheet was attached to a CP that had been removed for renovation (Fig. 14.15) and CP scan was used to inspect this area. The pole had cracks ranging in width from 0.3 to 0.15 mm. Although this preliminary experiment was suitable for confirming the method's principle, the attached bill-posting prevention sheet was not bonded to the pole. The scanning rate was 120 mm/s with data being acquired at intervals of 0.2 mm in the scanning direction. That rate will allow one CP to be inspected in about three minutes. A millimeter wave image was obtained after three scans (Fig. 14.16). Although the equipment must be calibrated, the imaging only requires the equipment to be kept tightly against the pole during the scan, so the operation is simple.

To allow a single person to perform the inspection, as shown in Fig. 14.16, the power supply and PC components of the system must be small and light, and image processing is essential for automatic crack detection from the acquired images. Therefore, NTT Access System Laboratories is currently developing an algorithm suitable for MMW image analysis.

These experiments used the Crack Scan automatic crack detection software that was subjected to hard disclosure commissioning to AIREC Engineering Corporation in 2008 for evaluation of detection characteristics. The red line in the bottom image of Fig. 14.17 indicates cracks detected from the Crack Scan automatic detection algorithm, confirming that 0.15-mm-wide cracks can be detected.
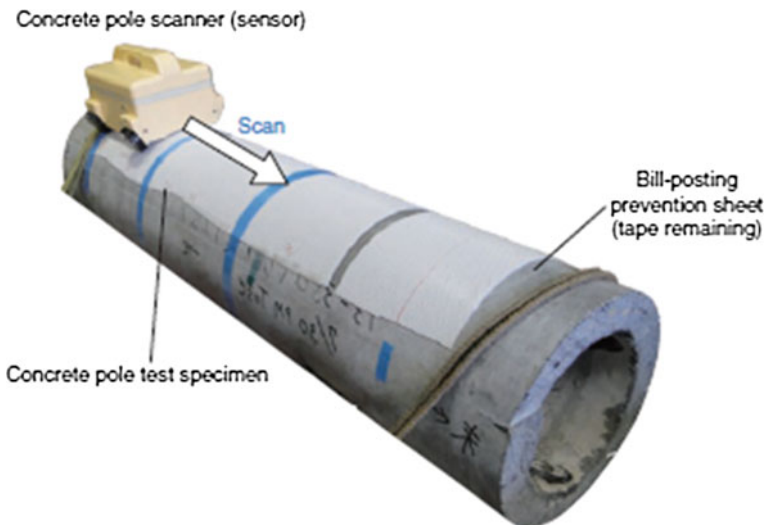
Fig. 14.15 Experimental setup
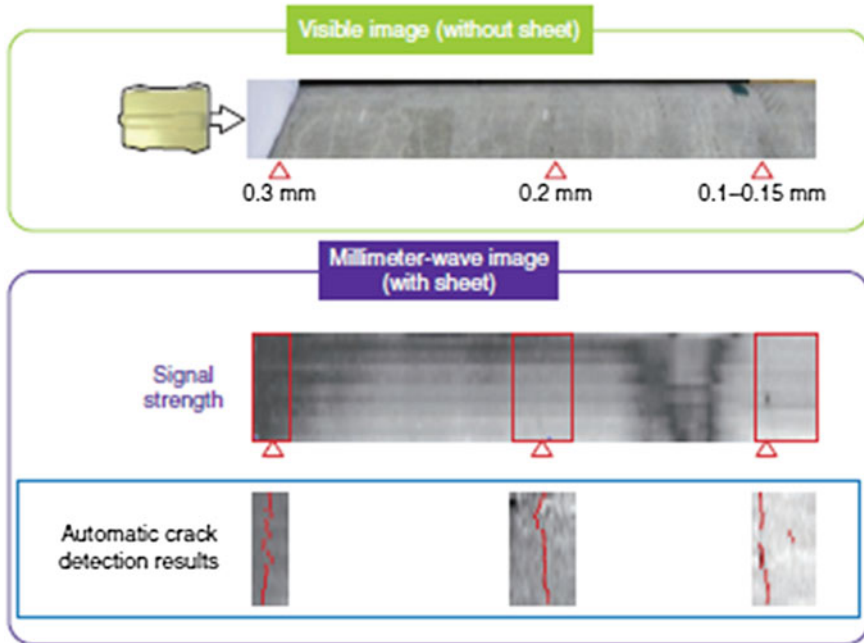
Fig. 14.16 Results of crack detection in test specimen
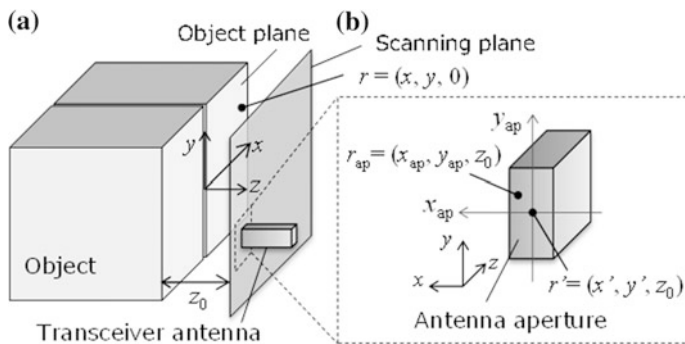
Fig. 14.17 (**a**) Position of the object, the center of the antenna aperture, and (**b**) antenna aperture in xyz coordinates

These results open up the prospect of developing portable equipment that can detect cracks that are at least 0.15 mm wide with a scanning time of about three minutes per pole.

14.3.3.4 Image Reconstruction for Crack Scan

To improve the spatial resolution in the radiative near-field region, Kojima and Mochizuki applied signal processing based on a holographic technique to MMW near-field imaging. A basic holographic algorithm has been proposed to detect concealed weapons and demonstrated a high-resolution 2D image reconstructed with the near-field data measured over the surface of a target at a single frequency [24]. With this technique, it is assumed with the image reconstruction algorithm that near-field waves radiating from the antenna and backscattering from the target point are closely approximated by a spherical wave and that a small-aperture antenna is suitable for radiating the spherical wave. In practical measurement, however, a directivity antenna such as a horn antenna and open-ended oversized waveguide probe are used to obtain high radiation efficiency. With the image reconstruction algorithm, the difference between the near-field radiation pattern of the antenna in the measurement and assumption causes degradation of the spatial resolution of images.

To overcome this issue, we developed an image reconstruction algorithm for improving the spatial resolution of MMW images. Our algorithm is based on deconvolution processing with the near-field radiation pattern from the antenna. We now describe the basic concept of our algorithm and present the reconstruction results from images of a 0.1-mm-wide crack on the surface of a concrete block measured at 76.5 GHz.

The configuration of the measurement model for verifying our algorithm is shown in Fig. 14.17. With our algorithm, the propagation equation should be well described, including the contribution of the near-field radiation pattern. The response of the transceiver, $u_s(r')$, can be written as a superposition of the response from all positions on the object, taking the effect of the antenna aperture into account:

$$u_s(r') = \iint_{\text{aper}} A(r_{\text{ap}}) \iint u_i(r - r')f(r)g(r_{\text{ap}} + r' - r)\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}S_{\text{ap}} \qquad (14.11)$$

where $f(r)$ is the object function indicating the dielectric distribution on the surface of the object, $g(r_{\text{ap}} + r' - r)$ is Green's function, which is written in three dimensions as

$$g(r_{\text{ap}} + r' - r) = \frac{\mathrm{e}^{-ik|r_{\text{ap}} + r' - r|}}{|r_{\text{ap}} + r' - r|}, \qquad (14.12)$$

and $u_i(r - r')$ is the incident electric field at the object surface:

$$u_i(r - r') = \iint_{\text{aper}} A(r_{\text{ap}}) \frac{\mathrm{e}^{-ik|r - (r_{\text{ap}} + r')|}}{|r - (r_{\text{ap}} + r')|} \mathrm{d}S_{\text{ap}}. \qquad (14.13)$$

Focusing attention on the near-field radiation in Eq. (14.11), we can change the order of the integral equation:

$$u_s(r') = \iint u_i(r - r')f(r) \iint_{\text{aper}} A(r_{ap})g(r_{ap} + r' - r)\mathrm{d}S_{ap}\,\mathrm{d}x\,\mathrm{d}y. \qquad (14.14)$$

By using Eq. (14.3) and the condition $g(r_{ap} + r' - r) = g(r - (r_{ap} + r'))$, Eq. (14.14) can be simplified as follows:

$$u_s(r') = \iint f(r)[u_i(r - r')]^2 \mathrm{d}x\mathrm{d}y. \qquad (14.15)$$

Thus, $u_s(r')$ is written as a convolution integral of $f(r)$ and the square of $u_i(r)$. We can solve the object function $f(r)$ by applying a 2D Fourier transform to both sides of Eq. (14.15):

$$FT_{2D}[u_s(r')] = FT_{2D}\left\{[u_i(r)]^2\right\}FT_{2D}[f(r)], \qquad (14.16)$$

where FT_{2D} indicates the 2D Fourier transform. Thus,

$$f(r) = FT_{2D}^{-1}\left[\frac{FT_{2D}[u_s(r')]}{FT_{2D}\left\{[u_i(r)]^2\right\}}\right], \qquad (14.17)$$

where FT_{2D}^{-1} indicates the inverse 2D Fourier transform. Therefore, if it is possible to estimate the incident electric field function, the object function can be solved. In the case in which the amplitude decay in Eq. (14.13) and the antenna aperture size are neglected, Eq. (14.14) corresponds to Eq. (14.17) [24–26], with which the beam pattern of the transceiver antenna that is spherical is assumed. Equation (14.13) above corrects not only the phase but also the amplitude so as to be the same object function all the time. Therefore, a higher resolution independent of the antenna near-field radiation pattern can be obtained by applying our algorithm.

The measurement system consists of an MMW radio frequency (RF) module [27], two-axis scanning stage with a controller, lock-in amplifier, analog-to-digital converter, and computer for controlling them (Fig. 14.18). The RF module generates MMW signals at 76.5 GHz, and the antenna radiates them into free space. The signals reflected by the sample are detected as in-phase and quadrature (IQ) signals and downconverted to an intermediate frequency (IF) of 100 kHz. The baseband circuit then converts them from analog to digital and sends them to the computer. The IQ signals are converted into the complex wave u_s:

$$u_s = I + jQ \qquad (14.18)$$
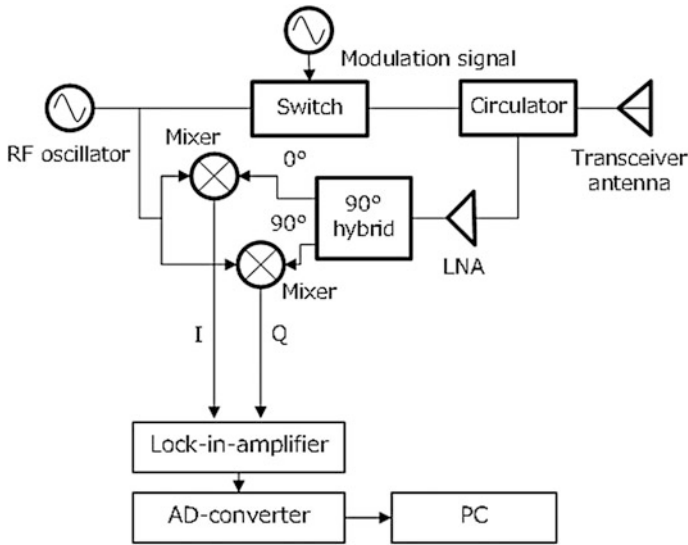
Fig. 14.18 Block diagram of imaging system

Kojima and Mochizuki used a concrete block with a 0.1-mm-wide crack and took MMW images of the crack using an open-ended waveguide antenna with an aperture size of 180×120 mm integrated in the Crack Scan. The distance between the surface of the concrete and antenna aperture was 10 mm. The scan used an aperture of 180×120 mm with a discretization of 2048 (x) $\times$ 256 (y). The data acquired were one-dimensionally processed along the x-axis in the same manner as the simulation. An (a) optical image, (b) raw MMW image, and (c) reconstructed image with the holographic algorithm, and (d) one with our algorithm are shown in Fig. 14.19. The image reconstructed with our algorithm was found to greatly improve in the contrast compared with the raw MMW image and exhibited better contrast than the image reconstructed with the holographic algorithm. Aspect ratios of the signal without and with the holographic algorithm and our algorithm were on average 4.4 ± 0.07, 2.4 ± 0.04, and 1.8 ± 0.05, respectively, in each image.

14.3.3.5 Future Prospects

For future work, we will test the system's practicality on actual poles with bill-posting prevention sheets bonded to them and evaluate the detection performance with our automatic crack detection software for MMW images of concrete poles. We will continue conducting R&D that will have wide applicability with highest priority on ease of operation and higher inspection efficiency while achieving practical detection performance suitable for equipment that can be used in actual concrete pole inspection.
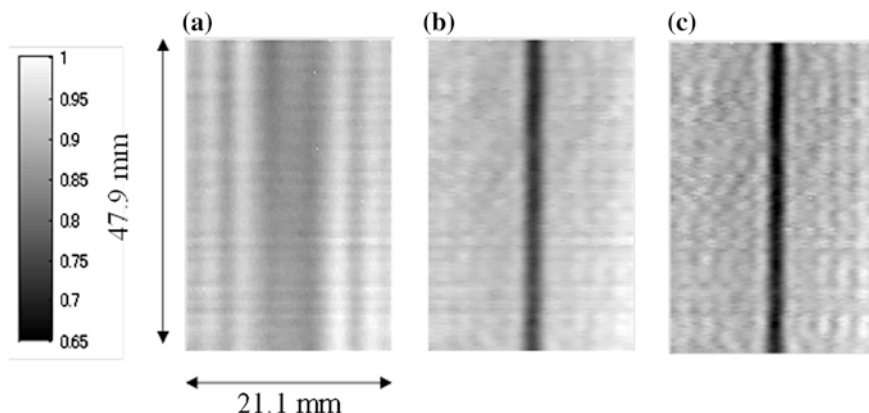
Fig. 14.19 Millimeter wave images without any algorithm (**a**) and images reconstructed by holographic (**b**) and our (**c**) algorithms

14.3.4   Image Processing for NDV of Concrete Structures

14.3.4.1   Crack Detection Problem

1. Importance of crack inspection for concrete structures

Decrepit concrete structures, such as expressways built a long time ago, have become a crucial concern. For maintaining them, crack detection in concrete surface is one of the most important roles in determining the deterioration of structures. When cracks appear on a concrete surface, water and air can penetrate the concrete and attack the reinforcing steel inside, causing the steel to corrode. As a result, the strength of the concrete structure decreases, as shown in Fig. 14.1. To prevent this sort of deterioration, it is important to detect and repair these surface cracks as early as possible.

In many cases, the width of cracks appearing on concrete surfaces is used as an indicator of the degree to which cracks require repair. Depending on the proposed standards of each country, there is some variance in the allowable crack width for structural durability. Roughly, the allowable crack width is often set at 0.3–0.4 mm in dry air and less than that in moist air; cracks of 0.2 mm or more in width are to be repaired with a filler or cover, or are reinforced with a steel anchor [28].

Therefore, current crack diagnosis methods are primarily visual. However, even for skilled workers, long hours of work make it difficult to maintain the concentration needed to prevent the overlooking of cracks less than a millimeter in width. When inspecting objects in high places, workers peer through binoculars rather than with the naked eye, which uses up their physical strength and concentration and increases the risk of overlooking cracks. Consequently, the use of machines to automate crack detection is desirable, and to make equipment acquisition and

fieldwork easier, the most plausible approach is to use digital cameras to take remote photographs of structures to automatically detect cracks from the images with image processing software.

2. **Difficulty in solving crack detection problems using image processing**

A number of algorithms have already been proposed for automatically detecting cracks from images of concrete wall faces, but these algorithms come with practicality issues for the following reasons [29, 30]. Usually, when trying to detect cracks by image processing, cracks are defined using feature vectors called edges, no matter what algorithm is used. Edges are the outlines that occur where there is a sharp change in shade between adjacent pixels. Because cracks appear as stringlike shadow lines in the image, edges are the only feature vectors indicative of cracks. These algorithms function effectively when the image shows only the concrete face, but they do not function well with images containing obstacles that are not part of the concrete face. The reason is that obstacles, such as windows, sashes, cables, and climbing gaffs, coexist in complicated ways with the concrete structure in actual field conditions, and the outline edges of these obstacles or in the background can be falsely detected as cracks. An example of this is shown in Fig. 14.20. This was the result when commercially available crack detection software was used to diagnose the field image of a CP that contained vertical cracks. Pixels filled with red are places judged to be cracks, but edges of the background, belt, plate, and so on have been falsely detected as cracks.

Overcoming this issue requires a process for removing the background and obstacles from the image before detecting cracks, but because the forms and characteristics of these are infinite in variety, a robust solution is not easy with current pattern recognition technology. In light of these facts, we created an algorithm that automatically removes obstacles from the concrete structure's image before crack detection processing.

14.3.4.2 Crack Detection Algorithm Using Image Processing and Neural Network Technologies

1. **Removing obstacles by texture analysis**

In the field, backgrounds and obstacles exist in numerous forms, and it is extremely difficult to develop an algorithm that uniquely defines their characteristics and directly eliminates them. Therefore, we solved this problem by combining the following two indirect approaches.

Step 1   Extract the structure and remove the background by inverting the extraction results.

Step 2   Extract the concrete face from the extracted structure domain and remove obstacles that cover the structure by inverting the extraction results.

Fig. 14.20 Results of performing diagnosis with crack detection software available on the market
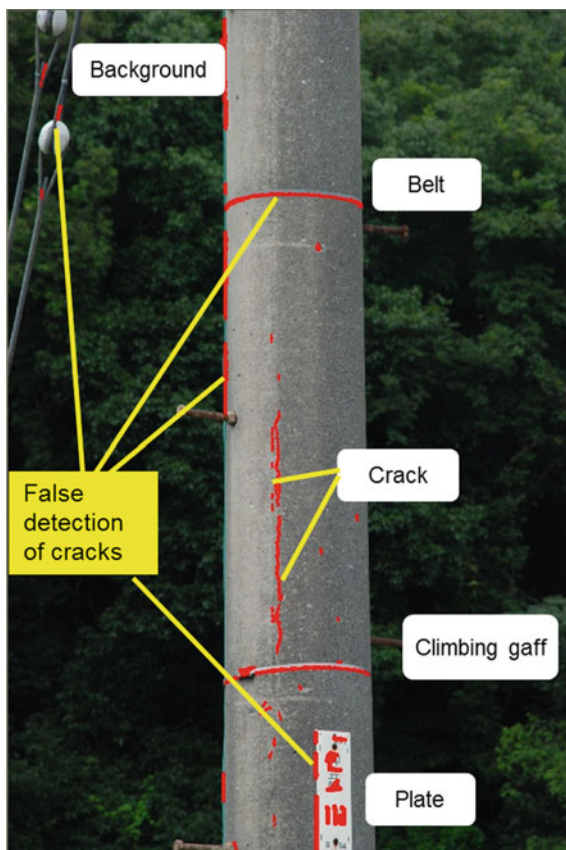
Taking an image of a CP as an example, we describe this proposed algorithm in detail below.

First, pixels with the color that seems to be that of the concrete face are extracted from the input image. Because concrete faces are ordinarily gray, the saturation of colors is quantified with the CIELAB color representation, as shown in Fig. 14.21, and pixels with saturation close to 0 are extracted. However, pixels in which the lightness is extremely high (pure white) or low (pure black) are removed. Figure 14.22 shows the color classification results: A "1" is assigned to each pixel with the color of the concrete face and a "0" to each pixel without the color of the concrete face (for the logic operation described later). Because this image has a green forest background, the CP domain was largely extracted by color classification, but it is often difficult to extract the CP from color classification alone, for example, if there are gray houses in the image. Consequently, robust extraction of CPs from various types of images requires texture data in addition to color data.

To extract pixels with texture like that of the concrete face, a standard deviation is calculated in relation to pixels' eight neighboring pixels, as defined in Fig. 14.23.

Fig. 14.21 CIELAB color system
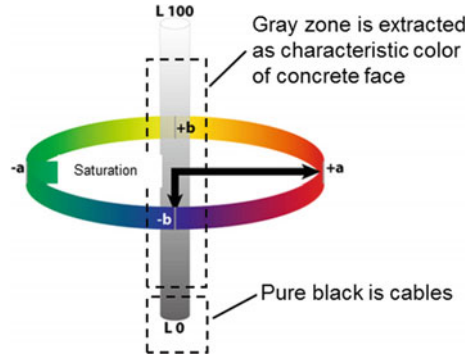
Fig. 14.22 Color classification results
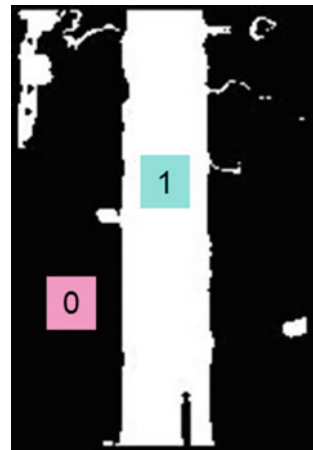
Fig. 14.23 Eight neighboring pixels
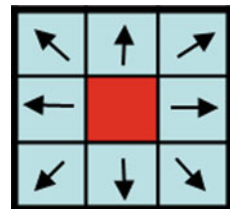
Results of calculating the standard deviation in relation to pixels' eight neighboring pixels in the input picture are given in Fig. 14.24. Generally, if the standard deviation of a concrete face is analyzed, it is at least 0.01 and less than 0.1, so pixels with a standard deviation in this range are assigned 1 and others assigned 0. The standard deviation calculation is addressed in the Handbook of Image Processing [31].

To extract the CP by combining color data and texture data, it is satisfactory to find the logical product of their output results. The logical product results are shown in Fig. 14.25. A vertical histogram of the results is given in Fig. 14.26. Concrete

Fig. 14.24 Texture classification results
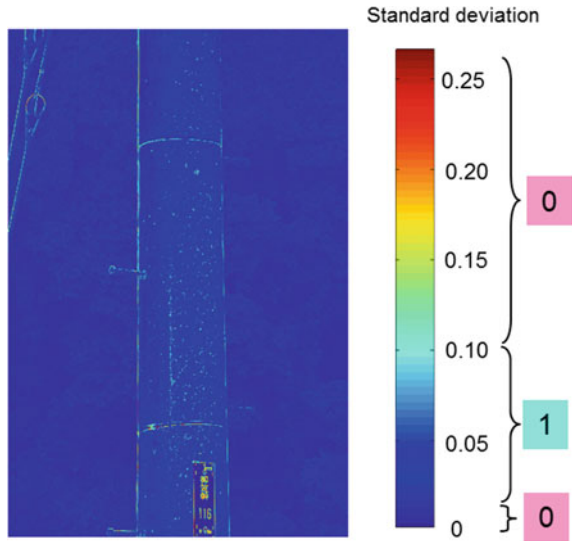
Fig. 14.25 Logical product of color and texture
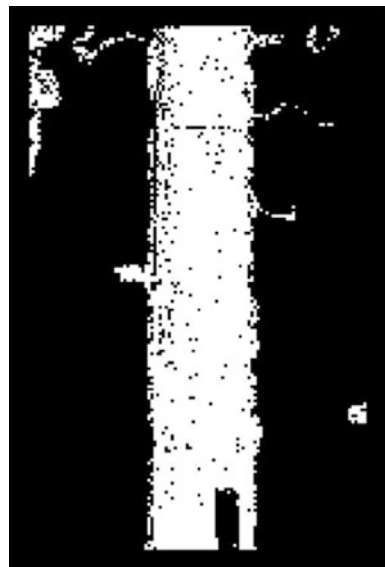
pole extraction as shown in Fig. 14.27 can be accomplished by performing threshold processing at the point that is half the peak value.

This algorithm can extract only the CP domain in front of the observation point even if other concrete structures appear in the background. This is because the camera is focused on the target in front of the observation point, so that concrete structures in the background are shown out of focus, and the judgment of the texture value removes these structures.

Fig. 14.26 Histogram of concrete candidate pixels
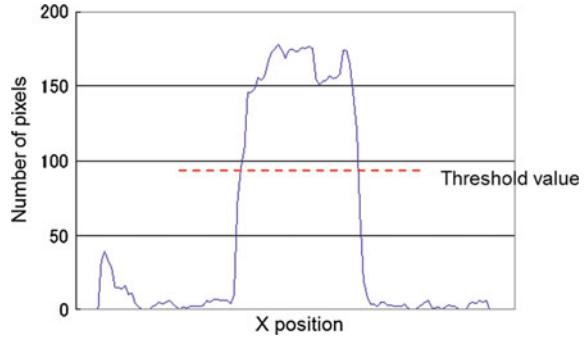
Fig. 14.27 CP extraction results

In the results shown in Fig. 14.27, the CP domain has been extracted from the input image without the background, but because the front of the CP is covered with obstacles such as belts, a plate, and climbing gaffs, it is not yet completely ready for crack detection. To extract only the concrete face without these obstacles, the edges of the obstacles are first extracted with the Canny edge filter. The results of edge extraction are shown in Fig. 14.28.

Next, the concrete pixels are extracted. Within the CP domain, the amount of texture S, which is called homogeneity as defined in Formula (14.20), is calculated for the 16 pixels (δ_1–δ_{16}), filled with light blue in Fig. 14.29, neighboring pixel A, which is filled with red and situated in the center. The i and j of Formula (14.19) are pixel luminance values; although they are ordinarily 256-bit grayscaled, they have been compressed to 8-bit grayscaled to shorten calculation time. The method for calculating the amount of texture is addressed in detail in Ref. [31].

Fig. 14.28 Edge extraction results

Fig. 14.29 Sixteen neighboring pixels

Fig. 14.30 Texture
comparison of concrete face
and area not part of concrete
face
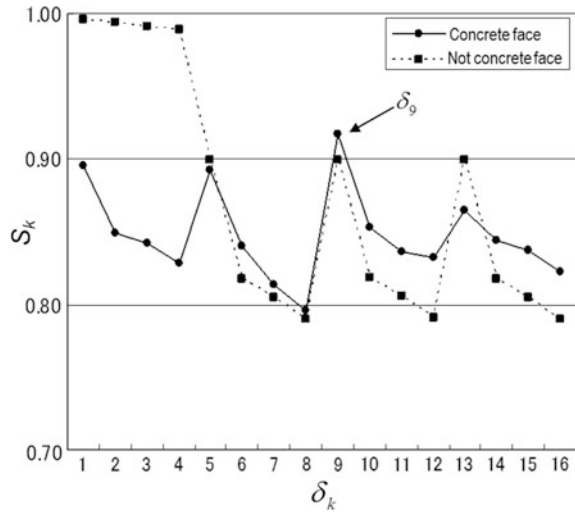
$$S_k = \sum_i \sum_j \frac{P_{\delta k}(i,j)}{1 + (i-j)^2} \tag{14.19}$$

Solving Formula (14.19) for the 16 neighboring pixels yields 16 amounts of texture S_k (k = 1, 2,…,16) as scalar values. The relationship between δ and S is shown in Fig. 14.30. The solid line shows the results calculated for a domain that is part of the concrete face, while the dotted line shows those for a domain that is not part of the concrete face. As Fig. 14.30 indicates, the concrete face is characterized by the fact that maximum peak S_9 is found at δ_9. Here, it is prescribed that "When k = 9, S_k is at its peak, and when the dispersion value of S_k is at least 0.0001, point A is a concrete pixel." To supplement this explanation, in domains with little roughness, the value of S is flat overall. Since it is possible that S_9 is the maximum by a small margin only by chance, the fact that S_9 is the maximum is not enough to define a feature vector of the concrete face, and consequently, the second condition that "the dispersion value of S_k is at least 0.0001" is necessary.

When a "pixel such that when k = 9, S_k is at its peak and the dispersion value of S_k is at least 0.0001" is found from within the CP domain, that pixel is taken as the starting point for filling in the CP, and the starting point is expanded with the domain enlargement method known as morphology dilation, as shown in Fig. 14.31. The expansion of the pixel is set to stop when it runs up against an extracted edge in an area where an obstacle edge has been extracted. In this manner, it is possible to extract just the concrete face from within the CP domain, without obstacles. Results of extracting the CP concrete face are shown in blue in Fig. 14.32.

Also, the reason that different texture calculation formulas were used for CP extraction and concrete face extraction was due to the difference in calculating cost.
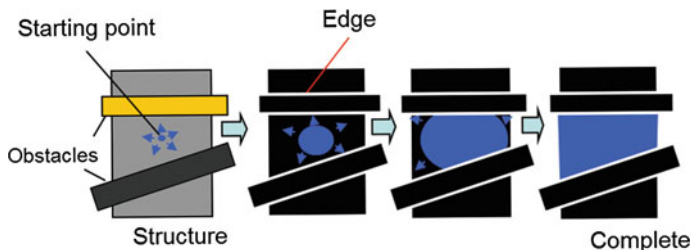
Fig. 14.31 Filling in concrete face by morphology

Fig. 14.32 Concrete face extraction results
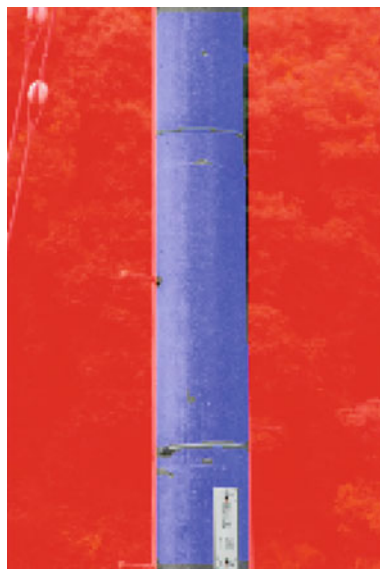
For CP extraction, the standard deviation for eight neighboring pixels, with a low calculating cost, is used because the entire range of the image is being processed, whereas for concrete face extraction, a co-occurrence matrix, which has a high calculating cost but also high precision, is used since it only searches a few times for starting points. Pixels not on the concrete face extracted by the process described above are part of the background or obstacles. Therefore, when an inversion process is performed, the background and obstacles are completely removed, as shown in Fig. 14.33.

2. **Crack extraction by recurrent neural network**

After removing obstacles, we extract cracks with a recurrent neural network [32–34]. The crack detection problem can be represented using an $x \times y$ neuron array, as shown in Fig. 14.34, where x and y are the size of pixels in the input image.

Fig. 14.33 Obstacle removal
results

Fig. 14.34 Expression of
problem by 2D neural
network
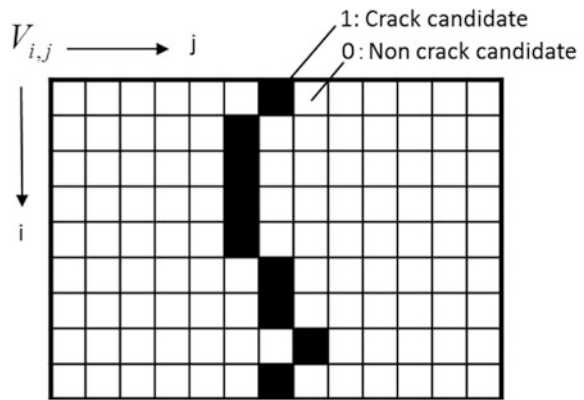
Here, V is neuron output (1: crack candidate; 0: not crack candidate) and U is
input energy. The search for a solution follows the following procedure.

(i) Initialize the input of neurons U with uniform random values.
(ii) Update the status of V according to the maximum neuron rule indicated in
Formula (14.21).

$$V_{xy}(t+1) = \begin{array}{ll} 1 & \text{if } U_{xy} = \max\lfloor U_{ky}; \forall : k \rfloor, \\ 0 & \text{otherwise} \end{array} \qquad (14.20)$$

(iii) For each neuron, update the value of U using the first-order Euler's method indicated in Formula (14.22).

$$U_{xy}(t+1) = U_{xy}(t) + \Delta U_{xy} \tag{14.21}$$

ΔU is given by Formulas (14.23)–(14.25).

$$\Delta U_{xy} = \alpha(t) \cdot p(x, y) - \beta(t) \cdot q(x, y) \tag{14.22}$$

$$p(x, y) = \sum_{m=x-M}^{x} \left\{ I_{m-1,y} - I_{m,y} \right\} + \sum_{m=x}^{x+M} \left\{ I_{m+1,y} - I_{m,y} \right\} \tag{14.23}$$
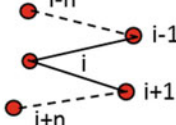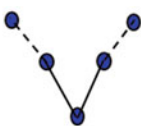
$$q(x, y) = \sum_{b=y-B}^{y-1} \| (x, y) - (a, b) \| + \sum_{b=y+1}^{y+B} \| (x, y) - (a, b) \|, \tag{14.24}$$

where $V\_\{ab\} = 1$

$$\alpha(t) = \alpha\_init - \{ (t-1) \cdot \alpha\_init \cdot \alpha\_grad \} \tag{14.25}$$

$$\beta(t) = \beta\_init + \{ (t-1) \cdot \beta\_init \cdot \beta\_grad \} \tag{14.26}$$

Here, I_{xy} represents the intensity of pixel (x, y), and $\| (x, y) - (a, b) \|$ represents the Euclidean distance between pixel (x, y) and pixel (a, b). As shown in Fig. 14.35, function $p(i, j)$ is the edge gradient, and the sharper the gradient of the edge is, the

	$p(i, j) \propto$ 2-D edge	$q(i, j) \propto$ Connected distance
Low possibility of crack	Edge gradient is shallow	Connected distance is long
High possibility of crack	Edge gradient is steep	Connected distance is short

• n: a parameter determining the length of cracks to be detected

Fig. 14.35 Two amount of characteristics in which crack is decided
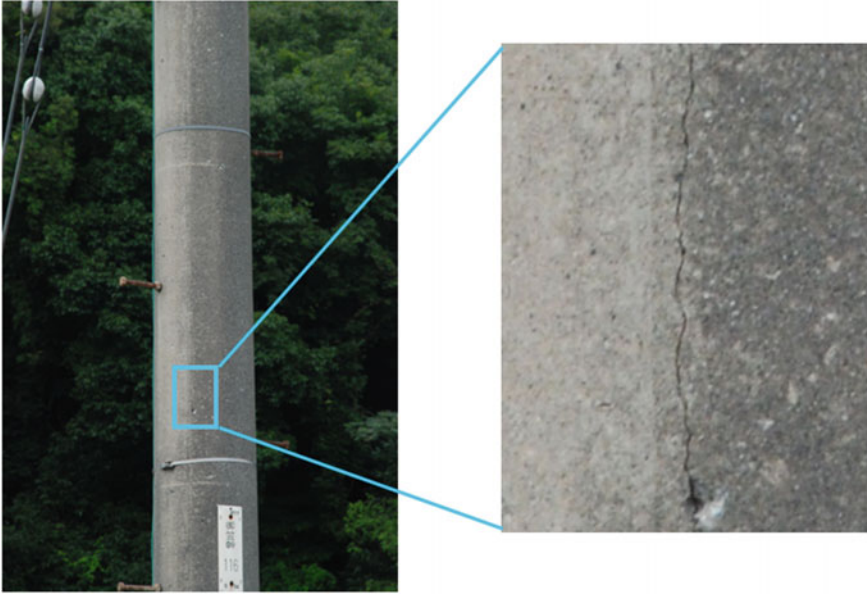
Fig. 14.36 Crack image of the concrete pole

greater the value returned. Function $q(i, j)$ represents the connectivity of a firing neuron at a 2D coordinate, and the longer the connection distance is, the greater the value returned. The edge gradient coefficient $\alpha(t)$ and connection distance coefficient $\beta(t)$ are given in Formulas (14.25) and (14.26).

(iv) Go to step (ii) until the status of V converges to equilibrium.

We clipped a crack area of the original image, as shown in Fig. 14.36, for simulation. The size of the crack image was 300×300 pixels with 8-bit grayscale, and the width of the crack was about 0.1 mm. In this simulation, we set each parameter: $t = 15$, $\alpha\_init = 500$, $\alpha\_grad = 0.07$, $\beta\_init = 1$, $\beta\_grad = 1.2$, $M = 2$, and $B = 30$, respectively. Figure 14.37 shows the result of simulation, where a vertical crack was successfully extracted. The computational time with C programming language and a Core (TM) 2 CPU was about 0.15 s.

14.3.4.3 Various Applications for Civil Engineering

This crack detection algorithm can be widely applied to common concrete structures other than CPs. For example, Fig. 14.38 shows the results of crack detection in a building wall image. In this case, our algorithm removed separator holes and edges' frameworks as obstacles and detected a vertical crack of 0.2 mm wide successfully. Figures 14.39 and 14.40 show the results of crack detection in a

Fig. 14.37 Simulation result of crack detection

Fig. 14.38 Crack detection of building image
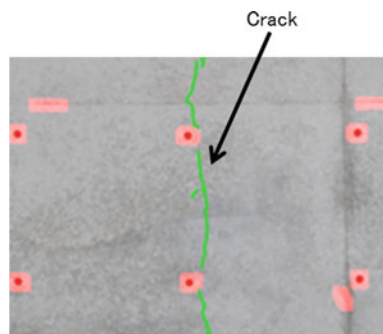
Fig. 14.39 Crack detection of manhole image
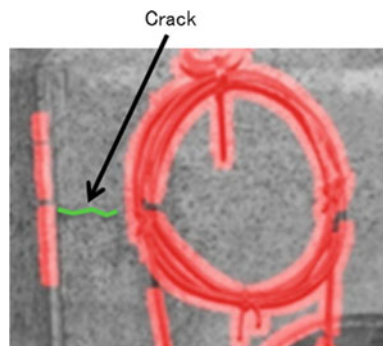
Fig. 14.40 Crack detection
of network tunnel image
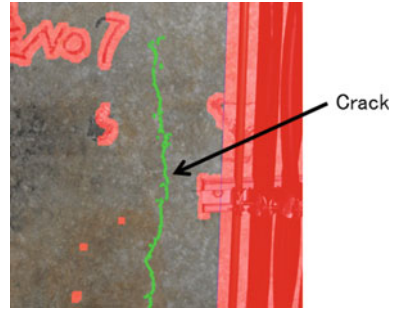
manhole image and a network tunnel image, respectively. In these cases, our algorithm removed some network cables on the wall in both images and hand-writing characters in Fig. 14.40 as obstacles, and extracts horizontal and vertical cracks of 0.2–0.3 mm width successfully.

14.3.4.4 Future Prospects

We proposed this image processing software in order not to overlook cracks. To further improve the accuracy of crack detection, we need to reveal the process of human visual perception and convert it to executable algorithms. Because the entire mechanism of human visual perception is really complex, trying to implement all of them on a computer is distant. Therefore, we should focus on essential and well-selected elements that are inevitable to resolve the crack detection problem. Our recurrent neural network model is nothing more than one of them. Another future work is to develop an administrative database for saving a huge amount of images. For checking the secular change in deteriorations on a database, we can carry out efficient maintenance of aged concrete structures.

14.4 Nondestructive Visualization of Wood Using EMW

14.4.1 Wood

14.4.1.1 Anatomical Features of Wood

1. Origin of wood as material

Wood is a material made of the plant cells of trees. The stem of the trees above the ground grows year after year forming concentric layer structures which are called annual ring. The woody cells are produced in a vascular cambium allocated in the peripheral zone of the stem. The activity of the cell production changes

Fig. 14.41 Three sections of wood
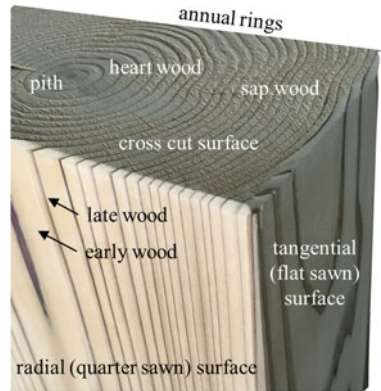
season to season, so that the concentric tone pattern of the annual ring appears in the cross section (Figure 14.41). Three types of sections are defined and referred often, cross (transverse) section, quarter-sawn (radial) surface, and flat-sawn (tangential) surface, respectively.

The cell produced in the vascular cambium grows for a while and then dies accompanied by the varnishing of the contents. Only the cell wall is left in this process, and a cavity appears in the cell. This cavity is used for transportation of water. The inner part of the concentric area is called heartwood, whereas the outer part is called sapwood.

2. Softwood and hardwood

There are two types of wood, softwood and hardwood, respectively. Both of them are used for buildings, furniture, musical instruments, and so on. Softwoods such as cypress, pine, spruce, fir, and cedar are the major species for wooden houses. Hardwoods such as oak, beech, birch, chestnut, maple, and walnut are the major species for furniture and musical instruments. Softwood consists of mainly the cells called tracheid that play the role of water transportation in wood (Fig. 14.42). At the same time, the aggregate of cells works as a load-bearing element. The wood cell is slender in the longitudinal direction. The length of the tracheid is about 3 mm and the diameter about 30 μm, and the thickness of the cell wall is about 3 μm. The hardwood structure is more complicated and consists of vessels, fibers, rays, etc.

3. Cell wall

The cell wall of wood has a multilayered structure made of a bundle of polycrystalline cellulose (microfibril). The bundle is reinforced by hemicellulose and lignin. The middle layer of the cell wall is oriented by 10–20° from the longitudinal axis of the cell and in a spiral manner (Fig. 14.43). This cell wall structure is the origin of the strength of wood, and the cell wall is very strong in longitudinal direction rather than in radial or tangential direction. The ratio of cellulose, hemicellulose, and lignin is about 45:30:25.
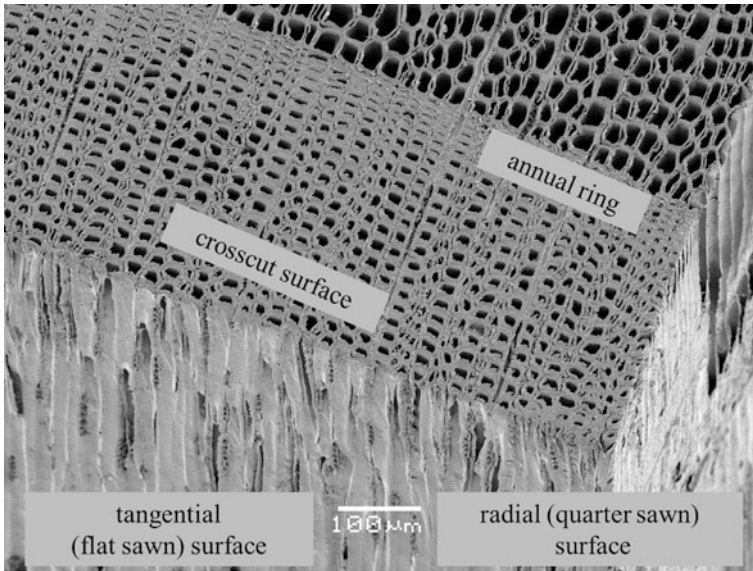
Fig. 14.42 Three sections of wood (softwood) observed by electron microscope

Fig. 14.43 Model structure of wood cell wall
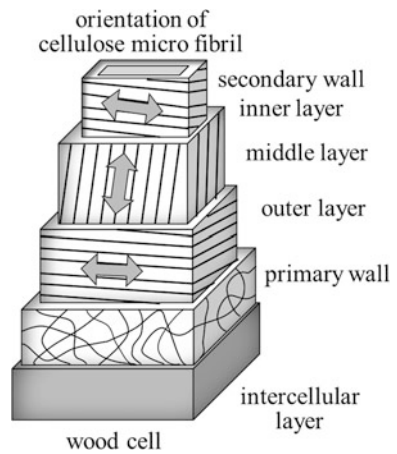
4. Anatomical features and defects

Wood grain that appears on the cut surface is influenced by anatomical features such as annual ring structures and may add to the ornamental value of wood. On the other hand, some anatomical features such as knots are considered as defects and may subtract from it a value.

14.4.1.2 Basic Properties of Wood

There are multiple parameters that express the physical, mechanical, and other properties of wood. Wood is the material made of plant cells, so these parameters are strongly influenced by anatomical or chemical characteristics of wood cells. These parameters are also influenced by moisture content and specific gravity.

1. Specific gravity and porous structure of wood

Specific gravity of the cell wall (substantial part of wood without cavities) is about 1.5, regardless of wood species. The apparent weight of wood depends on the ratio of the cavity in a unit volume, and this also causes the difference of elasticity or strength among wood species. Specific gravity is 0.35–0.5 for Japanese cypress, 0.4–0.5 for red pine, and 0.45–0.8 for zelkova wood. Specific gravity of paulownia known as light wood is 0.2–0.4, and the one of oak, known as heavy wood, is 0.8–0.9.

2. Moisture in wood

Wood is a hygroscopic material. It absorbs and desorbs water. The ratio of water in wood is expressed as moisture content. It is defined as the percent ratio of the weight of the water in wood to its oven dry weight (Fig. 14.44). The moisture content of fresh and green wood is very high, although it varies greatly in and between individual wood. It is higher in sapwood than in heartwood for softwood.

When wood is located under room temperature and relative humidity for a long time, it reaches a state, in which the amount of water absorption from the atmosphere becomes equivalent to the one of desorption, and the moisture content reaches a constant value. This state is called air dry, and the moisture content at this state is about 15 %, although it varies depending on the climate.

There are two states of water existent in wood, free water and bound water, respectively. The former is the water drop or condensed liquid water in the cell cavity. The latter is the absorbed water in the cell wall and bound to the cellulose structure by a weak bonding force. Wood under moisture content of about 28 % contains mainly bonded water, with the amount of free water increasing over this moisture content. This moisture content is called the fiber saturation point (FSP).
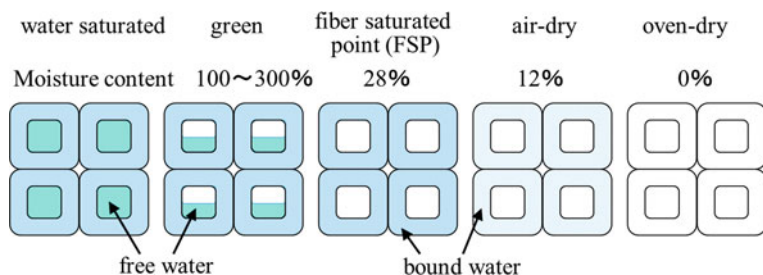
Fig. 14.44 State of water in wood and moisture content

It is the boundary of the state of water distribution in wood, and many physical and mechanical properties of wood change at this boundary FSP. For example, wood swells in absorption and shrinks in desorption. The ratio of this dimensional change is large under FSP and less over FSP. In addition, the ratio of this dimensional change shows anisotropy.

3. Thermal properties of wood

Wood has a porous structure and contains a fair amount of air in it, so that it shows a high thermal isolation effect. Wood of lower density shows a higher thermal isolation effect. The coefficient of thermal conductivity of wood is 0.1–0.13 kcal/m²h°C and is much lower than the one of concrete (1.4–1.6) or steel (46). In wooden constructions, there are many interfaces between wood and other materials such as metals, concrete, and stone that have high heat conductivity. Water condensation often occurs in this interface and is dependent on the temperature and relative humidity. The condensed water is absorbed into the wood, and it causes biodegradation.

4. Mechanical properties and strength of wood

The most popular parameters for elasticity of material are Young's modulus (modulus of elasticity). Most of the parameters of elasticity increase in accordance with specific gravity. However, as was explained previously, specific gravity of the cell wall is constant and about 1.5, so that the difference of elasticity among wood species is caused by the amount or distribution of cavity or wood substance in a unit volume.

Elasticity of wood is strongly influenced by other factors, such as anisotropy and moisture content. Elasticity of wood is different in major orientations, longitudinal (L), radial (R), and tangential (T), respectively. The order of Young's modulus (E) in these orientations is $E_L \gg E_R > E_T$, and the order for shear moduli (G) is $G_{LR} > G_{LT} > G_{RT}$, respectively.

Young's modulus in bending is about 90,000 kgf/cm² for Japanese cypress, 75 for Japanese cedar, 115 for red pine, and 120 for zelkova wood, respectively. Young's modulus in longitudinal direction is about 134,000 kgf/cm² for Japanese cypress, 75 for Japanese cedar, 120 for red pine and 105 for zelkova wood, respectively. Young's modulus in radial direction is of about a tenth of the one in the longitudinal and about twentieth in the tangential one.

Anisotropy in mechanical properties of wood comes mainly from the alignment of wood cells. In softwood, for example, an annual ring consisted of a pair of early wood of lower density grown in spring and late wood of higher density grown in summer (Fig. 14.45).

Parameters of mechanical properties decrease in accordance with moisture content in the range from dried status up to FSP. In this range, the bound water swells the cell wall and causes the decrease in the flocculation effect of the cell wall structure. The influence of moisture content over FSP is slight (Fig. 14.46) [35].

Strength of wood is expressed in various forms according to the external load applied, such as tensile, compression, bending, shear, torsion, and cleavage.

Fig. 14.45 Anisotropy of wood strength

longitudinal   radial   tangential
100   :   10   :   5
ratio of compression strength
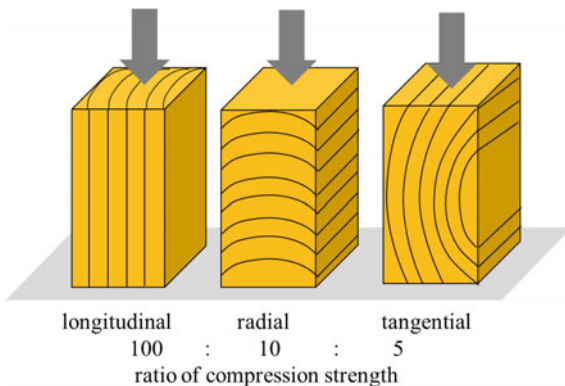
Fig. 14.46 Relationships between moisture content and strength of wood. *Note* modified from the literature by Markwardt et al. and by USDA [35]
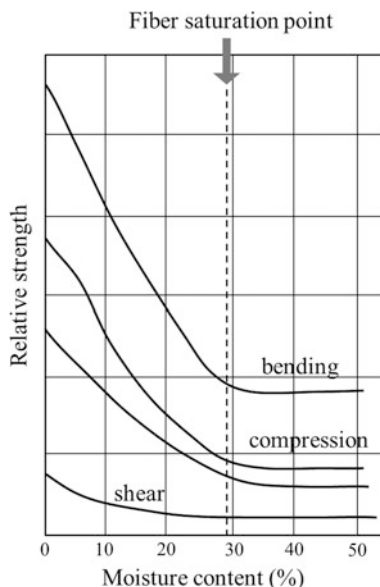
Normally, these strengths are evaluated in static loading. The longitudinal strength is normally the largest in tensile test, followed by bending and compression strengths; however, the order is influenced by anatomical features such as knots.

Among the ratios of the parameters of elasticity and strength to specific gravity, the elastic modulus in longitudinal direction of softwood is almost the same as the one for steel. This means that wood is strong given its lightweight.

14.4.1.3 Biodegradation of Wood

The major defects of wood and wooden constructions are as follows: combustibility, instability in dimensions caused by water absorption and desorption, and deterioration by attack of biological agents such as insects or fungi. In the practices of production and usage of wood and wooden architectures, quality control for these factors is essential and the nondestructive inspection technologies for quality evaluation are imperative. In this section, the outline of the degradation by biological agents called biodegradation is reviewed.

1. Decay caused by fungal attack (Figs. 14.47, 14.48, 14.49 and 14.50)

   Wood cells consist of natural polymers (polysaccharides) such as cellulose and hemicellulose. They are edible for fungi, that is, wood can be decomposed by the
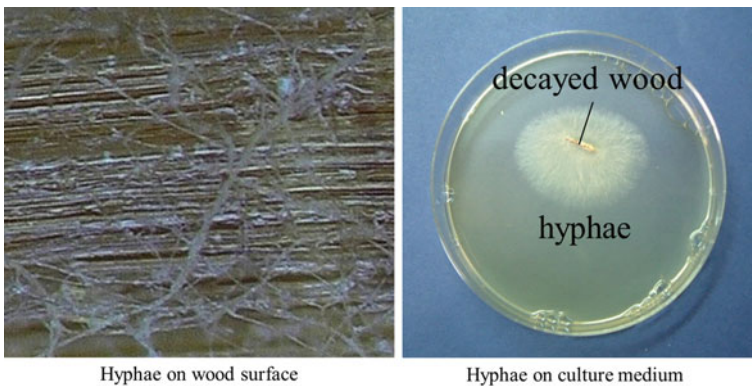
Fig. 14.47 Hyphae of wood-decaying fungi
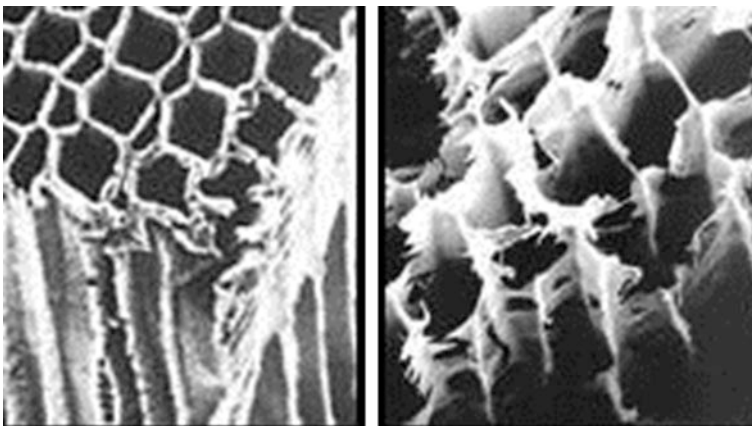
Fig. 14.48 Cell wall structure, sound (*left*) and heavily decayed (*right*). *Note* photographs by Yuji Imamura

Fig. 14.49 Wood decay in incipient stage (*left*) and progressed stage (*right*) featured by hyphae

Fig. 14.50 Wood decay in final stages

digestive activity of fungi such as basidiomycetes, and as a result, the density and strength decrease. Airborne spores attached to wood surfaces cause germination and division by water supplies under moderate climate conditions. The growth of hyphae will decompose the wood with enzymes, and the cell wall will be lost. The development of a fungal attack occurs when the wood is exposed to adequate temperature (0–50 °C), air (oxygen, say 20 % of wood volume), moisture (30–150 % moisture content), etc. However, the control of temperature or the removal of air is actually impossible, because the optimum conditions for these

factors are almost the same as for humans. The controllable factor left behind is the moisture content of wood. So long as the moisture content of wood is kept under FSP (fiber saturation point of about 28 %), wood will not deteriorate by fungal attack, because under these conditions, no free water exists in the cellular structure, with which the fungi can grow. On the other hand, a lack of fungal attack generates in wood on extremely high moisture content (higher than 150 % or water-saturated), because the fungi cannot be exposed to a sufficient amount of air under this condition. Therefore, it follows that in order to control fungal attack in wood and wooden buildings, it is essential to control the moisture in wood. In actual wooden buildings, the water that moistens the wooden structural members is water by rain that enters through the roof or wall, water from the soil under the floor, water leaked from facilities, and dew condensed in the structures. These waters are generated in hidden or sealed areas in the buildings and are often not detectable by ordinary inspection methods in its early stage. Therefore, nondestructive detection of moisture associated with fungal attack is an essential technology.

2. Insect attack

Like fungi, insects attack wood as well as living trees. Well-known insects that attack wooden products are termites and beetles. They invade dried wooden members of the building and make holes by chewing and biting. As a result, the strength of wood sharply decreases.

Termites are the most significant wood-attacking insects, with more than 2200 species. About 50 species of them attack wooden buildings, with distribution mainly from tropical areas up to warm zones including North America, Oceanian, and Asian countries where wooden houses are popular. There are two types of termites: subterranean termites living mainly in the ground or moist wood and dry wood termites living in dried wood (Fig. 14.51). Termites are so-called social insects, meaning they may form colonies of 10,000 up to million worker and soldier termites. The subterranean termites live in a nest in the ground and they form a network of galleries through which they move and seek wood as bait. Once they

Fig. 14.51 Example of subterranean termite

soldier   worker

Fig. 14.52 Termite galleries developed on wood under floor level

Fig. 14.53 Wood attacked by termites

reach a wooden building, they crawl up the structure and start attacking. Termites are adverse to light or airflow, so they crawl up in dark, hidden, or enclosed structures and materials. Usually, termites attack without notice (Figs. 14.52, 14.53 and 14.54) and it is difficult to detect the attack by visual inspection, especially at its early stage. Dry wood termites start their attack by swarming. Nymphs, after pairing, invade the wood and form a new colony. Their attack is often seen in doors, window frames, and furniture; however, they also inhabit structural members of wood, and it is difficult to discover their attack. Termites are small insects whose length is about several millimeters; however, their colonies are hugely populated and they attack wooden constructions secretly and severely. Termites attack by repeatedly chewing and bite off small fractions of wood. They attack other materials such as plastic, paper, bamboo, and concrete.

Fig. 14.54 Termite attacks developing in wooden wall

Other types of dry wood-attacking insects include deathwatch beetles and powder post beetles. Adults of these insects invade wood and yield eggs. Larvae from the eggs attack the wood by chewing and biting. Like termites, it is difficult to detect the attack of beetles. It is important to have nondestructive inspection tools to detect and control the attack by termites and beetles in wood and wooden buildings.

14.4.2 Electrical Properties of Wood

The most important electrical properties of wood are electrical conductivity, the dielectric constant, and the loss tangent, respectively. Electric conductivity is the measure of electric current that flows in a material, when the material is placed under a given voltage gradient. The dielectric constant is the measure of the polarization or the electric potential energy induced and stored in non-conducting material, when it is placed under an alternative electric field. The loss tangent is the ratio of energy lost in the repetition of polarization. These parameters are specific for every material; however, in the case of wood, they depend on density, moisture content and anatomical features and anisotropy of the properties. There have been multiple developments for wood drying and adhesion using electrical properties. Among these, the measuring apparatus of moisture content using electrostatic capacitance is a well-known application.

14.4.2.1   Electric Conductivity

In a conductive material such as metal, electric current flows by the movement of free electrons. On the other hand, in non-conductive materials such as wood, the activity of the ions associated with the polar band in the cell wall and the ions derived from non-organic components in non-crystal area plays an important role. This is related to the mechanism of polarization discussed afterward.

Electric conductivity λ is given by the following formula:

$$\log \lambda = B - C/\varepsilon, \tag{14.27}$$

where B is a constant related to Faraday's constant, C a constant related to energy of ionization, and ε dielectric constant, respectively.

The electric conductivity of wood is influenced by moisture content, temperature, density, and anatomical features. The most important factor is moisture content (Fig. 14.55) [36]. Electric resistance expressed as inverse of conductivity decreases linearly with moisture content from oven dry up to several percent and decreases exponentially until FSP and decreases slightly over FSP.

Specific resistance decreases with increases in temperature, and this tendency is noticeable in lower moisture content. The influence of temperature can be attributed to the ionization in the cell. The influence of the wood species and density is not significant. Electric resistance is subjected by the cell alignment. Specific resistance in the direction normal to fiber direction is larger by 2.3–4.5 than the one in fiber direction for softwood, and by 2.5–8.0 for hardwood. The resistance in the tangential direction is slightly larger than that in the radial one. The chemical components cellulose and lignin in the wood cell wall are the origin of the electric resistance; however, the effect of the water-soluble electrolyte is larger for the moisture content over FSP. When the wood contains preservatives or fire retardant that contains metallic components and electrolytes, the resistance is larger. The properties of the resistance to alternative current (impedance) under several thousand Hz are the same for direct current, and the influence of polarization is much larger over the frequency.
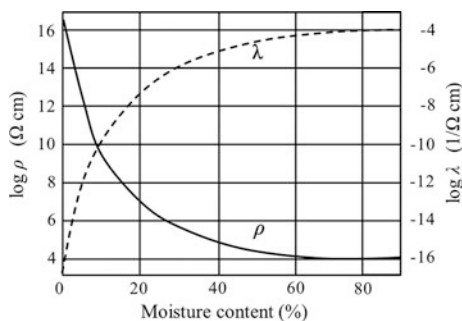
Fig. 14.55 Relationships between moisture content and λ and ρ for 10 European species. *Note* modified from the literature by Lin and by FPS [36]

14.4.2.2 Dielectric Constant

Dielectric constant as the measure of feasibility of polarization of a material is normally expressed as the relative value of the material's dielectric constant to the value for vacuum state. Experimentally, it is the amount of electric quantity accumulated in a unit volume of wood by polarization, when the wood is placed under voltage gradient between a pair of electrodes. Dielectric constant for air takes 1, for water 81, for oven-dried wood 2–4, and 6.7 for dried cellulose.

The dielectric constant of wood is influenced largely by moisture content and density (Fig. 14.56) [37] and increases in accordance with moisture content. The increase becomes larger in the range of 15 up to 20 %, and it reaches about 10 at FSP of about 30 %. This increase depends on the density. For lower density (specific gravity) wood, the constant is 3 up to 5 at FSP, whereas it is about 8 for wood of specific gravity about 0.5.

In the moisture content over FSP, it increases largely due to the free water in the wood cell, and this is more apparent for high-density wood. The high-density wood contains more substances that take part of polarization in comparison, so that the dielectric constant becomes larger. The dielectric constant becomes larger for higher temperature. However, the increase is several percent for the temperature increase of 10 °C and is negligible for normal room temperature. The dielectric properties show anisotropy and are influenced by anatomical features of wood, especially for higher frequency. The dielectric constant decreases as the frequency increases; however, this change is strongly influenced by moisture content (Fig. 14.57) [38]. As is previously explained, it increases as moisture content for the same frequency.

14.4.2.3 Loss Tangent, $\tan \delta$

An ideal dielectric material releases all the accumulated energy when the electric field is eliminated by the movement of electric quantity through the connected outer

Fig. 14.56 Relationships between moisture content and dielectric constant at 1 MHz for various wood densities. *Note* modified from the literature by Uemura and FFPRI, Japan [37]
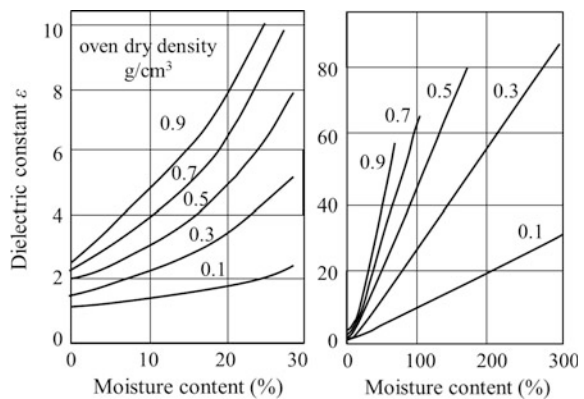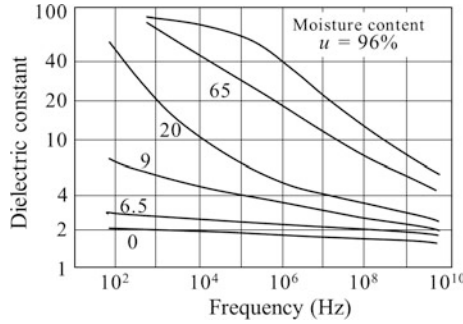
Fig. 14.57 Dependence of dielectric constant on frequency and moisture content. *Note* for spruce, in tangential direction at 20 °C, modified from the literature by Trapp et al. [38]

circuit. However, actual dielectric materials lose energy partly in the form of heat. Power factor is the ratio of lost energy to the accumulated. Experimentally, it is related to the deviation of the electric current to the voltage gradient expressed by deviation angle (phase angle) φ, when a wood is sandwiched by a pair of electrodes and placed under an alternative electric field. The phase angle φ is the delay of current to the voltage gradient when wood is modeled electrically as a parallel circuit consisting of a resistance and a capacitance and placed under an alternative voltage gradient. The value of $\cos \varphi$ is called dielectric power factor and it varies between 0 and 1.

The angle δ as complementary angle of the angle φ is called loss angle, and $\tan \delta$ is called loss tangent. When the value of δ is much smaller than φ, loss tangent $\tan \delta$ can be approximated as $\cos \varphi$. Principally, $\tan \delta$ is given by the following formula:

$$\tan \delta = 1/2 \ \pi f CR = 1.8 \times 10^{12}/\varepsilon\rho f \qquad (14.28)$$

where f denotes frequency, C capacitance, R resistance, and ρ density, respectively.

Loss tangent is influenced by frequency, moisture content, density, and fiber orientation. It increases in accordance with moisture content but becomes much larger for higher density of wood (Fig. 14.58) [38].

Loss tangent change is a complicated manner in relation to frequency and to moisture content (Fig. 14.59) [39]. It takes a maximum in the frequency range between 10^7 and 10^8 Hz for the wood of moisture content under 3 %. The maximum value increases in accordance with moisture content over 6.5 %; however, loss tangent takes higher values for lower frequency in accordance with moisture content. In higher frequencies, the contribution of the orientation polarization in non-crystalline area of cellulose, hemicellulose, and lignin should be important. For lower frequencies, on the other hand, the interface polarization of conductive ions should be important. Loss tangent increases as the temperature increases; however, the influence of temperature is small between 10 °C and 30 °C.

Fig. 14.58 Relationships
between frequency and loss
tangent tan δ for various
moisture contents. *Note* see
Fig. 14.57
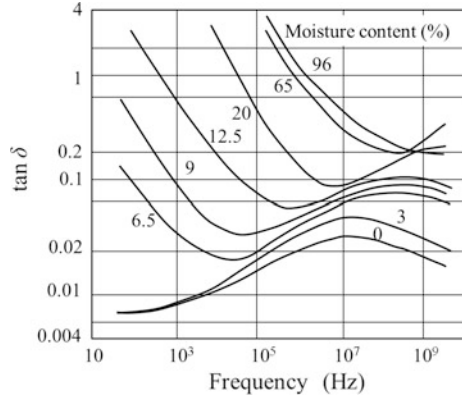
Fig. 14.59 Relationships
between moisture content and
loss tangent tan δ at 5 MHz.
Note modified from the
literature [39]
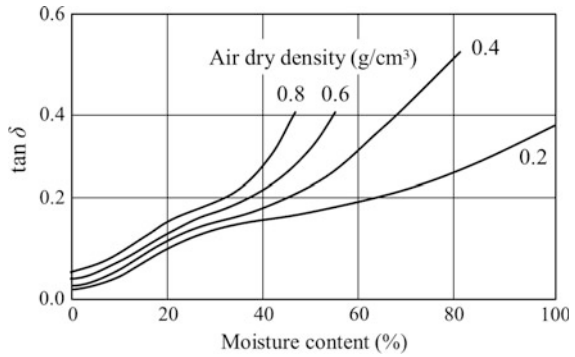
14.4.3 NDT of Wood Using Dielectric Properties of Wood

The most well-known apparatus and technique is the moisture meter. The dependency of electric resistance and capacitance on moisture content is applied to portable moisture meters. The moisture meter using capacitance can be applied to measure the moisture content of a relatively wide range and they evaluate the average moisture content for the area from wood surface down to a deeper area. However, the capacitance depends not only on the moisture content, but also on the density of wood, so that density adjustment is required in the measuring.

Another example is the pulse radar apparatus using microwave of the frequency range 1 to several GHz for the inspection of wood. Its aim is to detect inner holes, defects, or deterioration by catching the wave that reflected at the electromagnetic boundaries in wood. However, the difference in dielectric properties between wood and air is relatively small, so that the reflection signal is not always significant. In addition, it is difficult to obtain a higher resolution due to the frequency range and system configuration.

A radar apparatus to detect the activity of wood-attacking termites using microwave has been developed. Bodies of termites crawling in wood have high water content, creating electromagnetic boundaries in a dry wood. By observing the fluctuation of the standing wave, the activity of the termites can be evaluated.

The situation of practical NDT inspections of wooden buildings is different from the one in production lines of factories or in laboratories. In practical situations, more attention must be paid to the performance of the apparatus; for example, it should be nondestructive and non-touching, has real-time-processing affording results in 2D or 3D images, has enough measuring capacity, possesses higher resolution, and is reliable, portable, robust, and economic. Many of the apparatus previously developed use the so-called active method; that is, some physical energy such as vibration, heat, X-ray, or electromagnetic waves is given to the wood or wooden construction to be tested, then the reaction of the energy is detected in a transmitting or in a reflecting system, and the properties and state of the object are evaluated from the change in the signals. It is important that the practical apparatus should be of a reflection system, because the pair of emitter and receiver of the transmitting system cannot be applied to most of the practical situations.

In this chapter, NDT technologies for wood and wooden constructions using electromagnetic wave are introduced. In this technology, the frequency of the apparatus plays an important role. Microwave whose frequency is lower than 30 GHz transmits relatively effectively and can be applied to larger objects. It has an economical advantage; however, the resolution of imaging is relatively low for wood inspection. On the other hand, teraherz wave (0.3–3 THz) can afford extremely high resolution; however, due to the high attenuation of the wave in the material, the analysis is limited to the subsurface area, and in general, the measurement is expensive to conduct. Millimeter wave (30–300 GHz) (MMW) whose wavelength is about 10–1 mm can afford an appropriate resolution for imaging of wood, because it transmits into the wood to some extent.

14.4.4 Reaction of Wood to Electromagnetic Wave

14.4.4.1 Transmission of Millimeter Wave Through Wood

Figure 14.60 shows the setup for the millimeter wave transmitter and receiver system. A Gunn diode was used to generate the millimeter wave at an average output power of 10 mW at 100 GHz. The output signal from the diode was modulated by a PIN diode. A millimeter wave polarized linearly in vertical direction and radiated from a horn antenna is transformed into a parallel beam using a Teflon plano-convex lens. The beam was irradiated on the wood, and the transmitted wave is focused onto a sensor with another plano-convex lens. The sensor receives the electric field component parallel to the vertical direction. The received wave signal is processed in a lock-in amplifier, and the magnitude of the electric field is obtained as a voltage.
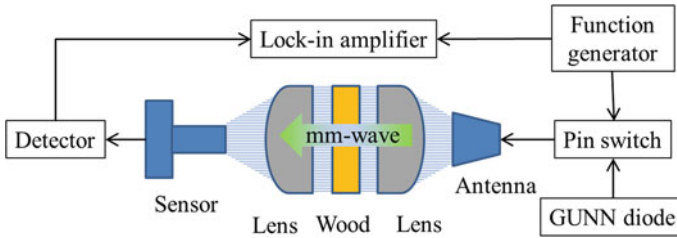
Fig. 14.60 Experimental setup of transmitting system
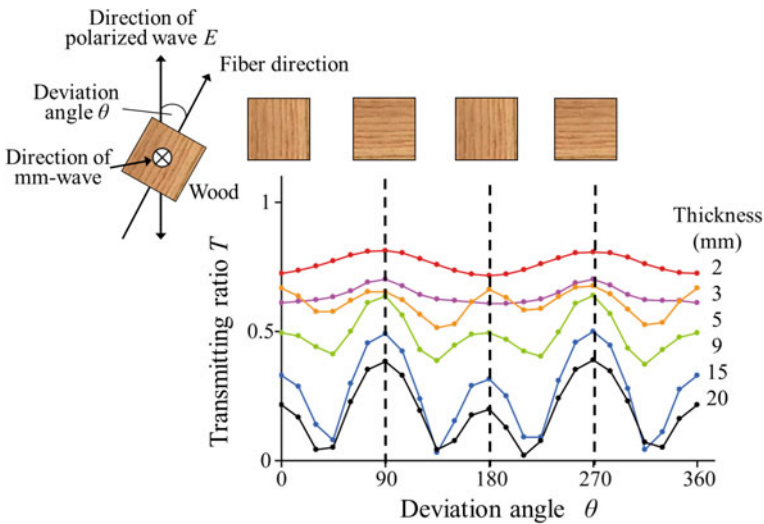
Fig. 14.61 Dependence of millimeter wave transmission on fiber direction

1. Transmission of MMW through wood board

Figure 14.61 shows the relationship between the fiber deviation angle θ and the magnitude of the transmitted MMW for flat-sawn boards of oven-dried cypress wood. The vertical axis denotes the ratio of transmission T (the ratio of the detected amplitude with and without wood $|E_{detect}|/|E_{in}|$. The transmission ratio is about 0.7 for a board of 2 mm in thickness and it decreases as the thickness increases. The transmission ratio changes periodically with the fiber deviation angle. For boards of 3–5 mm in thickness, the ratio takes the maximum for the fiber deviation angles of 90 and 270° and the minimum of 0 and 180°. For thicker boards, the number of the angle to give maximum and minimum values becomes larger. For a board of 20 mm in thickness, the minimum ratio is almost 0. There is no difference in these tendencies between flat-sawn and quarter-sawn boards.

2. Curve fitting based on theory of the transmission

The dependency of transmission ratio (attenuation) on the fiber direction can be attributed to the anisotropy of dielectric properties of wood. The mechanism of the anisotropy should be discussed in relation to the chemical components or anatomical features of wood. The main component of the cell wall is polycrystalline cellulose oriented regularly in fiber direction $(-C_6H_{12}O_4-)_n$. The polarization moment takes a maximum when the direction of electric field becomes parallel to the fiber direction (the orientation of cellulose polycrystalline) that is when $\theta = 0$ and $180°$. As a result, the incidence of MMW attenuates and the transmitting ratio decreases. The cyclic change of transmission ratio to the fiber orientation can be attributed to the birefringent behavior of wood, and wood behaves as a retardation plate to MMW. This behavior is influenced by wavelength, wood thickness, and the anisotropy of phase shift of transmitting wave. No dependence of fiber direction on the transmitting ratio was found for the wood-based materials, such as plywood, particle board, MDF, and hardboard whose fiber orientation is random.

When wood is irradiated by MMW, attenuation and phase shift of the transmitting wave occur in addition to the reflection on the front and back surface. The reflection is featured by reflection coefficient t. On the other hand, the attenuation is featured by attenuation coefficient α, and the phase shift is featured by phase shift coefficient β. The attenuation and the phase shift are anisotropic, and coefficients α and β are separated into α_x and β_x in fiber direction and α_y and β_y in the direction normal to fiber direction, respectively. As a result, the electric field generated by the transmitting MMW changes in accordance with the deviation angle of the fiber direction to the electric field.

The transmitting ratio is expressed as follows:

$$T = t_{AW}t_{WA}\exp\left(-\frac{1}{2}\alpha_x d\right)\left|\cos^2\theta + \exp\left\{-\frac{1}{2}(\alpha_y - \alpha_x)d\right\}\exp\{-j(\beta_y - \beta_x)d\}\sin^2\theta\right|$$

(14.29)

where t_{AW} and t_{WA} denote the reflection coefficient at front or back surface and d thickness of the wood, respectively.

This formula consists of three members and they are about reflection $(t_{AW} \times t_{WA})$, about attenuation in the material $(\exp(-1/2\alpha_x d))$, and about phase shift in absolute symbol, respectively. The member phase shift is related to the fiber deviation angle and wood thickness.

Furthermore, the formula takes the forms when $\theta = 0°$

$$\ln T = -\frac{1}{2}\alpha_x d + \ln t_{AW}t_{WA}$$

(14.30)

and when $\theta = 90°$

Fig. 14.62 Relationships
between specimen thickness
and transmitting ratio. *Note*
modified from the literature
by Tanaka [40]
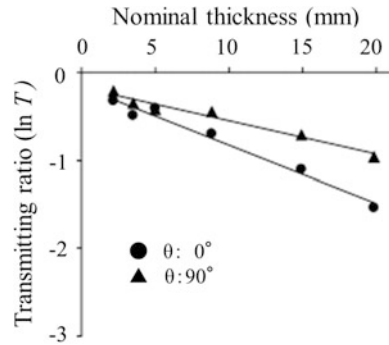
$$\ln T = -\frac{1}{2}\alpha_y d + \ln t_{AW}t_{WA} \tag{14.31}$$

The logarithm of the transmission ratio $\ln T$ is linear to the thickness d. The gradient corresponds to the attenuation coefficient at the given fiber deviation angle and the intercept the coefficient of reflection (Fig. 14.62) [40]. From the linear regression formula, the value of attenuation coefficient in the fiber direction is obtained as $\alpha_x = 0.13$ (/mm), the one in the normal direction as $\alpha_y = 0.08$ (/mm), and the reflection coefficient of 0.85.

3. The influence of moisture content

Figure 14.63 shows the dependence of the transmitting ratio on the fiber deviation angle of the test board at different moisture contents. The transmitting ratio decreases as the moisture content increases, and the dependency on the fiber direction becomes less significant. The moisture in wood absorbs MMW. The anisotropic factor expressed as the ratio of transmission ratios, along and normal to fiber direction, increases in accordance with in moisture content and takes a
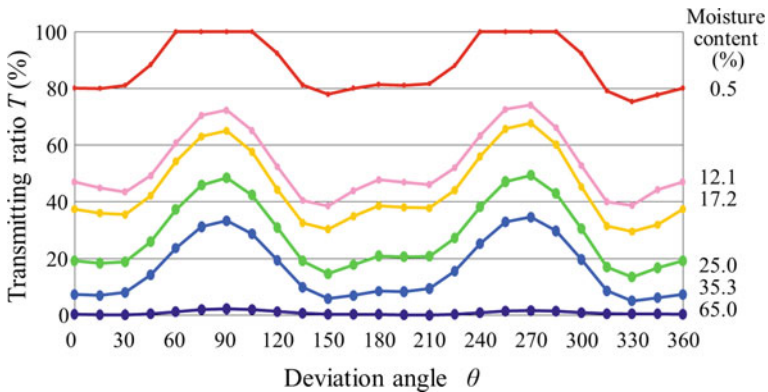
Fig. 14.63 Dependence of transmission on moisture content of Japanese cypress

Fig. 14.64 Dependence of anisotropy in transmittance on moisture content
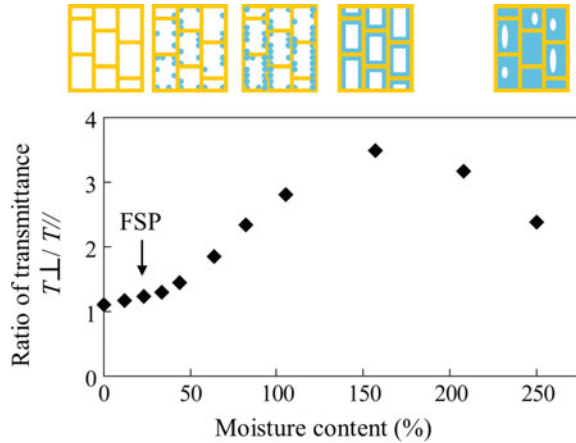
maximum value at about 150 %, and again decrease for higher moisture content as shown in Fig. 14.64. One possible reason is as follows: In moisture range under 150 %, water molecules are adsorbed to the cell wall and they are allocated as if the profile of the cell wall structure was enhanced, and as a result, the anisotropy of the transmitting ratio is enhanced. On the other hand, in the case of higher moisture content, the cell cavities are filled with water molecules and their allocation in fiber direction is diminished, and the factor of anisotropy becomes smaller. The coefficient of the attenuation of MMW in wood becomes larger as the density for both directions, along and normal to fiber direction, respectively. However, the attenuation can be different in wood species of the same density, and this suggests that anatomical features influence attenuation. More detailed discussion on MMW in the transmission in wood can be found in literatures by Tanaka et al. [41, 42].

14.4.4.2 Reflection of MMW from Wood

The behavior of millimeter wave in a reflection detection system was also investigated using a monostatic system of millimeter wave of 94 or 100 GHz (Figs. 14.65 and 14.66) [43, 44]. This apparatus is monostatic using a circulator. A Gunn diode was used to generate the millimeter wave at an average output power of 10 mW at 100 GHz. The output signal from the diode was modulated by a PIN diode, polarized and irradiated to the surface of wood through a circulator and waveguide probe whose cross section is square of 3 by 2 mm and 50 mm long. The reflected component is received by the same waveguide and sent through circulator to a Schottky diode for square law detection. The detected signal is sent to a lock-in amplifier, and the magnitude of the reflection is measured. This measurement is conducted by scanning of the wood specimen attached to a XY-table to get the 2D distribution of reflected amplitude. The distance between the aperture of the
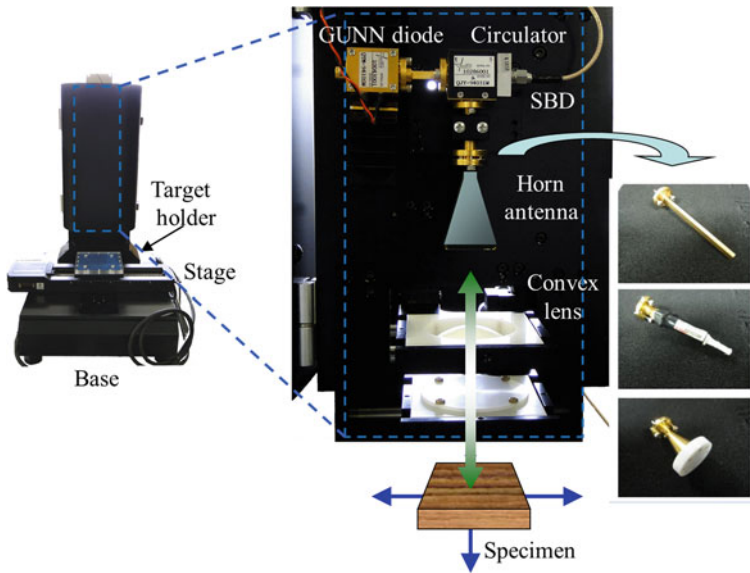
Fig. 14.65 Experimental setup for reflection imaging using 100 GHz millimeter wave [43, 44]
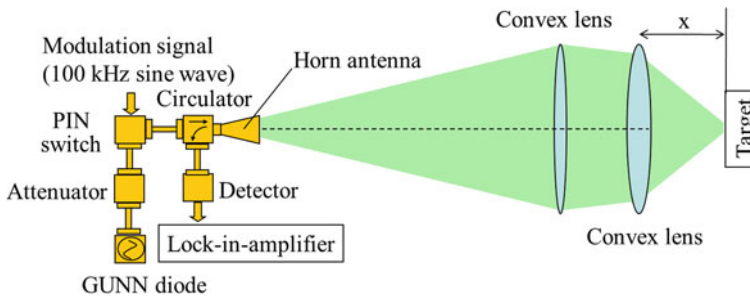
Fig. 14.66 System configurations

waveguide and the wood can be adjusted by 1 mm, and the direction of fiber can be set parallel or normal to polarization direction.

Imaging of surface and subsurface of finished wood was conducted. Figure 14.67 shows an example of image taken by scanning of wood block using reflection imaging by a monostatic system. In the scanning of wood surface finished with Japanese lacquer (Fig. 14.67a), the pattern of a wood grain orientation under the lacquer layer was clearly recognized (Fig. 14.67b). To clarify the penetration depth of millimeter wave into wood specimen, another test was conducted to detect the layer of aluminum foil inserted between a pair of cut wood blocks (Fig. 14.67c). In the reflection image, a characteristic stripe pattern is recognized near the left side of the image (Fig. 14.67d). This result was caused by the change in the amplitude of
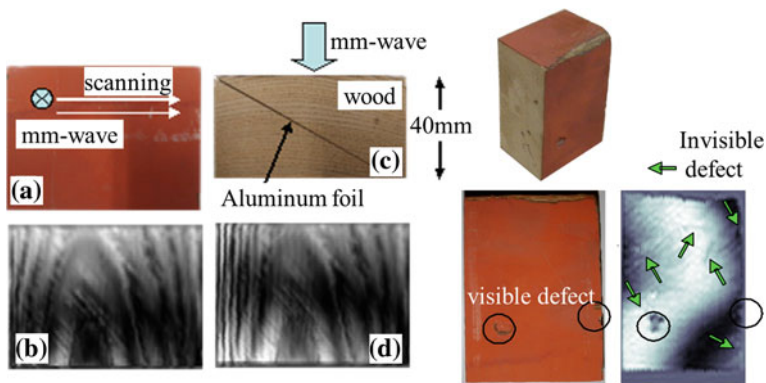
Fig. 14.67 Example of scanning using reflection system [43, 44]

the wave that is radiated to the wood, penetrated into the wood and reflected at the foil. The stripe pattern can be attributed to the interference of the penetrating wave with the reflected one. The existence of the inserted foil can be recognized at a depth of about 15–18 mm at the deepest case for the tested 5 wood species.

Figure 14.67 right shows an example of a reflection image of a small wood specimen taken from a post in a Buddhist temple (left). It was attacked by beetles and contains several holes. Some visible defects are also seen on the surface. The surface of the block was scanned by a millimeter wave apparatus. On the reflection of millimeter wave image, several characteristic shaded areas (dot line circles) are seen and they suggest the existence of inner defects.

The distribution of inner defects was also evaluated by a microfocus X-ray CT apparatus (SMX-160CT-SV3-S, SHIMADZU Corp.). This was compared with the MMW image. Characteristic changes can be seen in the X-ray image (left) corresponding to the shaded areas (dot line circles) suggesting inner defects in the MMW image. However, some other unique shaded areas can be also seen in the X-ray image. They can be attributed to the holes and galleries made by beetles at deeper positions, and they could not be detected by the millimeter wave imaging.

Another feasibility of reflection system is the detection of wood-attacking insects in the wood. Figure 14.68 [45] shows a time-domain signal of millimeter wave detected by reflection imaging system when a millimeter wave is irradiated on the surface of a wood specimen behind which a different number of living worker termites are put in a bottle. Millimeter wave is reflected at the electromagnetic boundary between wood and the insect, and this causes characteristic fluctuation of the amplitude. The frequency of the fluctuation depends on the number of termites inhabited. These results show that millimeter wave technology can be a useful nondestructive monitoring tool of activity of biological agents.

Fig. 14.68 Detection of wood-attacking insects inhabiting wood
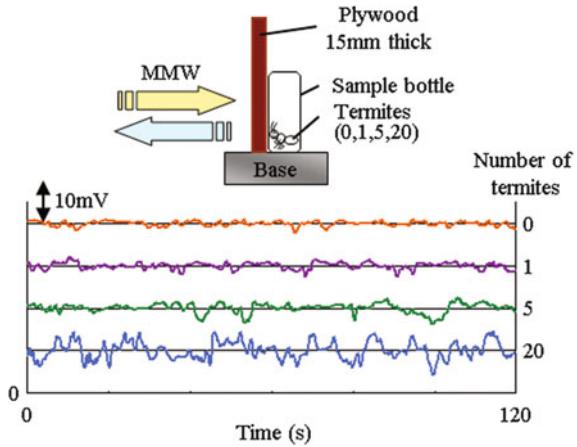
14.4.5 Wall Scanner for Diagnoses of Wooden Houses

14.4.5.1 Technical Background

Nondestructive inspection is one of the essential technologies for the structural health assessment of timber structures. Its aim is to analyze the structural property and other attributes, such as durability. The authors have recently developed a small radar apparatus for the nondestructive evaluation of wooden walls typically found in Japanese wooden family houses [46]. Two objectives were set in the development of this tool: to visualize the inner structure of the wall, giving us important information related to the evaluation of anti-earthquake resistance of the house, and to see the biodegradation or moisture distribution of inner wood related to the durability of the house (Figure 14.69) [46].

14.4.5.2 Devices [46]

The first developed experimental apparatus is a small scanner equipped with a horn antenna for sending and receiving of the electromagnetic wave attached to a vertical XY-stage of 400 by 400 mm of working area, a signal processor for generation and detection of electromagnetic wave, a controller for the XY-stage, and a small personal computer for the control of the measuring system (Figs. 14.70 and 14.71).

The apparatus works according to FMCW (frequency-modulated continuous wave) radar technology. The electromagnetic wave has a sweep range of 10–20 GHz emitted from the antenna and is irradiated to the wooden wall to be investigated. A part of the wave is reflected at the wall surface, but the rest of it transmits the wall. The transmitting components reflect partly at electromagnetic boundaries in the wall such as interfaces between wooden parts and the metal parts,
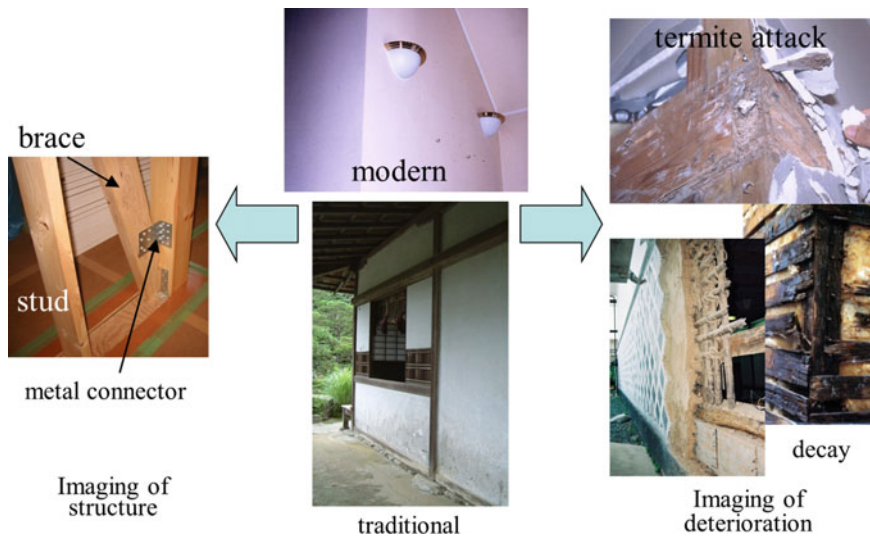
Fig. 14.69 Target walls for NDT using FMCW technique

Fig. 14.70 Scanning devices
with antenna
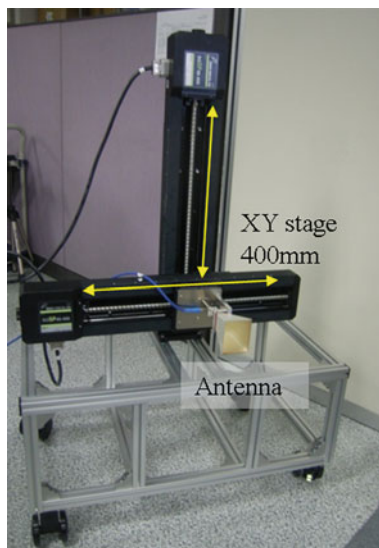
between inner cavities and substantial parts such as wood or metal, and between moist areas and dried areas. The reflected wave components are received by the antenna and detected by the signal processor.

In the signal processor, a beat signal is composed by synthesizing the source signal and the detected one. By analyzing the beat signal, the intensity and the

Fig. 14.71 Experimental setup of FMCW radar system to get time-domain signal

location of the reflection are obtained as the amplitude and the delay of peaks in the time-domain signal. By scanning of the antenna in the vertical plane using XY-stage and taking the data with an interval, the distribution of the intensity of the reflected signal in any vertical plane in the wall is obtained. At the same time, the change in the intensity of the reflected signal along the wall thickness at any point in the vertical plane is also obtained. By scanning, the time-domain signal can be displayed as 2D or 3D images, and it is possible to recognize the shape, the size, and the allocation of the members in the wall.

14.4.5.3 Performance of Wall Scanner for Structure Analysis [46]

Figure 14.72 shows an example of time-domain signal detected for a simple model wall structure consisting of two pieces of plywood of 12 mm in thickness and set at a distance of 95 mm. The time-domain signal can be obtained by the inverse FFT of the beat signal from the detector.

The horizontal axis denotes time, but actually it denotes the distance in the direction of wall thickness from the antenna. Two peaks are detected and they are the reflection of the two boards. The position of each peak corresponds to the location of the board, and the peak value changes in accordance with the dielectric constants of the board. The low peak found between the two peaks can be attributed to the multireflection of the wave in the system and it seems like a ghost.

Fig. 14.72 Time-domain signal showing feasibility of detection of objects hidden in a model wall [46]
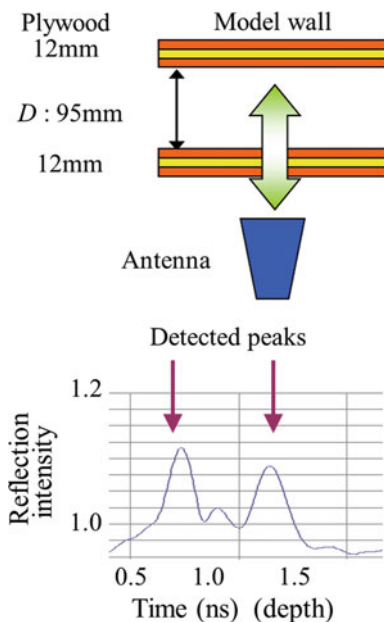
Figure 14.73 shows similar results except when a thin metal plate is attached to the backside of the board. The electromagnetic wave reflects at a boundary of dielectric properties. Larger reflection occurs at the boundary of the larger difference in dielectric properties. At the surface of metal, it reflects larger, and as a result, the peak in the detected signal becomes higher.

The performance of the apparatus was evaluated using a model wooden wall consisting of posts, beams, and a stud, and covered with a piece of plywood or a plaster board whose thickness is 12 mm (Fig. 14.74). The wooden members are assembled by metal plate connectors, bolts, and coarse thread screws. The posts and beams are made of laminated lumber whose cross-sectional dimension is 120 mm in square, and the stud made of solid wood is 45 mm in thickness. The wooden members are all made of dried softwood. The structure is a simplified model of the wall structure frequently found in modern Japanese wooden family houses constructed according to the posts and beams method.

This wall was scanned, and the results are shown in a contour map of the amplitude of the detected signals in a section at a distance from the antenna. The red area corresponds to the corner metal plate assembled in the wall, and the shape, the size, and the allocation of the plate can clearly be recognized. The wooden construction members such as studs, posts, and beams were also recognized, although the amplitude of the reflected signal was smaller than the one from the metallic parts. The volt of 12 mm in diameter in the post was also detected, so long as it is allocated at a depth less than 40 mm from the surface.

These results show that the apparatus is able to provide us important structural information hidden in the wall with a covering board. We have applied this method
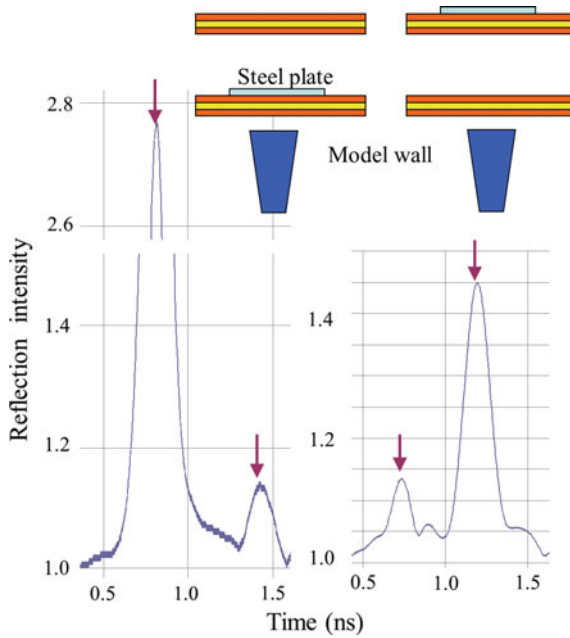
Fig. 14.73 Time-domain signal from model wooden wall
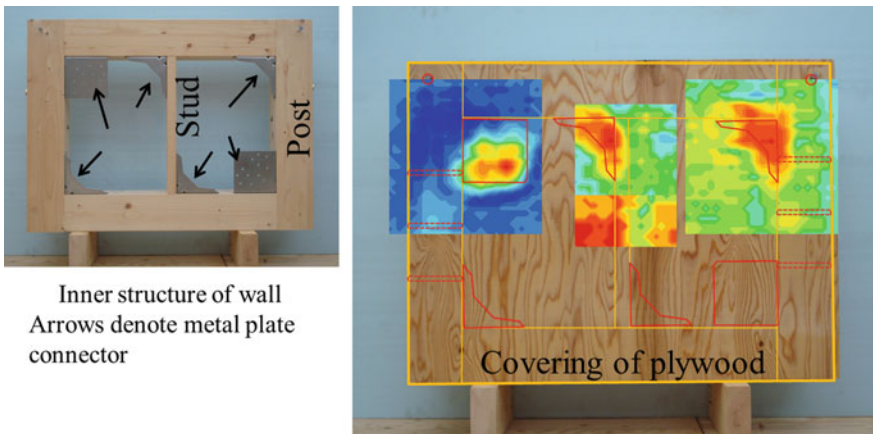
Fig. 14.74 Two-dimensional image of scanning of sample wooden wall

also to the Japanese traditional clay wall, and it was confirmed that the network made of bamboo strips (lath) set in the clay wall for the reinforcement was roughly recognized.
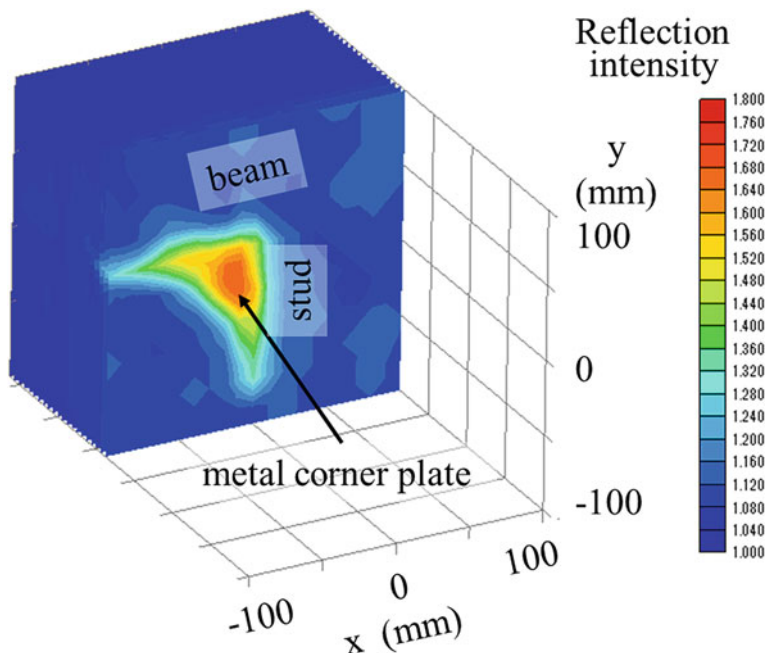
Fig. 14.75 Three-dimensional image of wall structure estimated by multiplying 2D distribution of reflection intensity

Figure 14.75 shows a distribution of the intensity of the reflected wave in a vertical section in the wall. The shape of the corner plate can be recognized easily. This distribution can be obtained at any layer of different distances from the antenna. As a result, the wall structure can be evaluated three-dimensionally. This is similar to the one obtained by computer tomogram (CT) technique using an X-ray apparatus. But the largest difference between them is that CT works in a transmitting system, whereas the wall scanner works in a reflection system. As already discussed, it is absolutely necessary to have a device of reflection system for practical NDT for buildings and other types of structures.

14.4.5.4   Performance of Wall Scanner for Biodegradation

When the wooden members in the wall are deteriorated biologically, for example, when they are attacked by termites or decayed by fungal attack, wood substance disappears and is accompanied by a decrease in strength, and this causes a change in the detected signal. More importantly, the biodegradation of wood is closely associated with the moisture in the wall. This moisture would be normally derived by water leaking through the covering of the wall or the condensation at the thermal boundaries in the wall structure. On the other hand, the electromagnetic properties
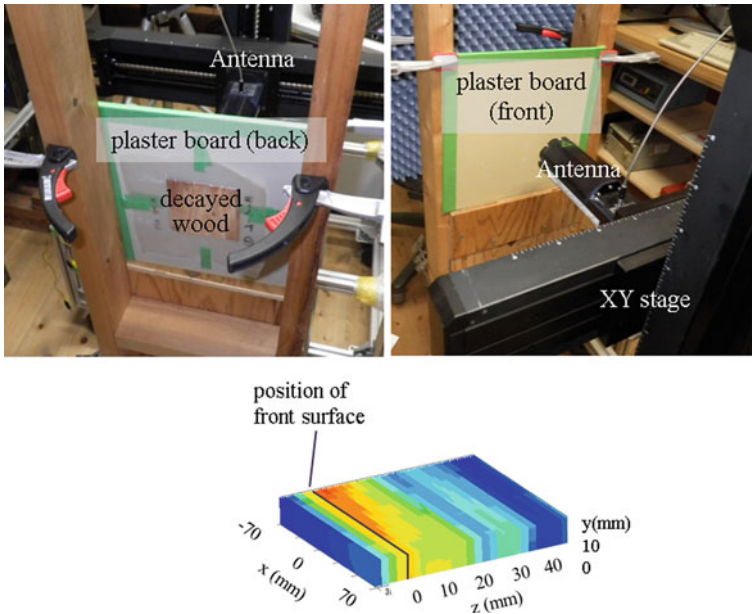
Fig. 14.76 Detection of moist wood hidden behind surface board

of wood vary by the moisture content of wood, so that it can be expected that by detecting the area of higher moisture content by the change in the reflecting electromagnetic wave, the possibility of biodegradation in the wooden members of the wall can be evaluated.

To confirm this idea, decayed wood plate of different moisture contents were prepared. They were sandwiched between a pair of polypropylene films and attached to the backside of a wall covering board. The board was scanned from the front side by the developed apparatus (Fig. 14.76). At the position of decayed wood plate of higher moisture content, significantly higher amplitudes of the reflected signal from the board were detected, although the peak amplitudes were lower than the ones from a metal plate attached to the back side of the board. The layered 2D images of the scanning show the distribution of higher peak amplitude corresponding to the moist and decayed wood plate and can be clearly recognized (Fig. 14.77). Under this condition, the area of moisture content is higher than the FSP and can be significantly detected, at which fungal activity in wood becomes higher.

Detection of decayed wood behind a sound covering board was also attempted. Artificial decayed wood of thin Japanese larch and Japanese fir boards exposed to fungal attack for 2–12 weeks (weight loss 2–50 %) was attached to the backside of a wall covering board, and this was scanned from the front side. It was confirmed that the decayed specimen in wet state was clearly recognized. However, the dried one was not always detected successfully. The result was that the electromagnetic wave was reflected at the boundary between the plaster board and the moist wood.
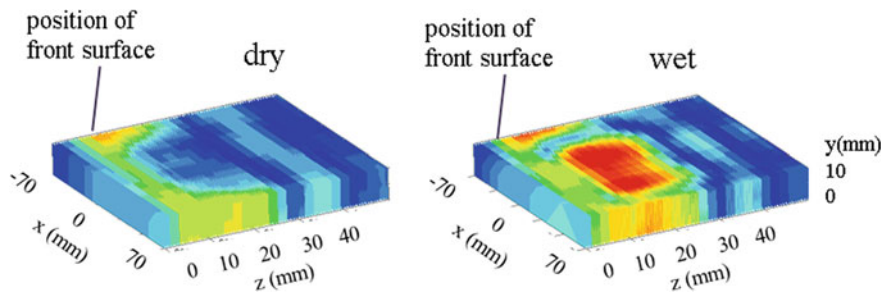
Fig. 14.77 Visualization of dry wood (*left*) and moist-decayed wood (*right*) hidden behind surface board

14.4.5.5 Development of a Portable Wall Scanner "Wood Wall Viewer"

Based on basic studies, a small radar apparatus was developed (Fig. 14.78) [46]. The apparatus consists of a small device and a personal computer. The principle of the detection is the same as the previous experimental one, but without the XY-stage and controller for scanning. The device is an all-in one type involving an antenna, electric circuits for sweep generation and detection of the electromagnetic wave, a data processing unit, and a battery. It is convenient to use for detection at one spot on a wall by attaching the device by hand, but it can be scrolled on the wall manually. The data are sent to a PC through USB interface and postprocessed by the developed control software. The detected signal is displayed on PC in the form of real-time time-domain signal in the manner of FIFO processing. By adjusting the analysis condition and by eliminating the signal caught for base or standard conditions, the target objects can be detected more precisely.

Fig. 14.78 "Wood wall viewer" using FMCW radar system
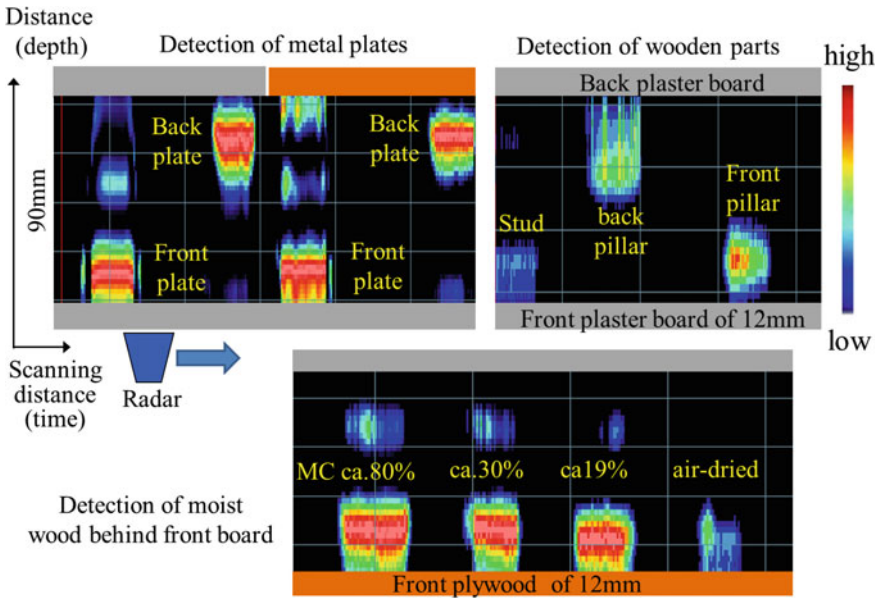
Fig. 14.79 Results of scanning using "Wood wall viewer"

Figure 14.79 shows some results of detection for a model wall consisting of two boards, face and back, set apart by 90 mm. The wall was scanned from the face side. A metal plate attached to the backside of the face board or to the front side of the back board is detected clearly. Studs of a smaller dimension located in the wall were also detected. A wood specimen of different moisture contents attached to the back side of the face board was also recognized. The intensity of the reflection increased in accordance with the moisture contents. On the other hand, some ghost peaks caused by the multiple reflections in the wall were detected.

14.4.6 Automatic Recognition of NDT Data Using Pattern Recognition

14.4.6.1 Introduction

The data obtained by NDT are in the form of signal, and 2D or 3D images and they often contain not only the information related to the target objects but also the influence of other factors or noise. For the proper interpretation of the data, the experience and knowledge of skilled inspectors are required. In this section, a technology of pattern recognition applied to an automatic interpretation of image data and to the recognition of defects in wood is introduced.

14.4.6.2   RCE Network

The technology of pattern recognition is an example of information process that recognizes an object, and actually, it is a flow (algorithm) of information process often expressed in the form of computer program. To recognize the object is to decide the category to which it belongs. For this purpose, a series of data measured from the object called "pattern" expressed in the form of vector is inputted into the program for pattern recognition and the program gives us the category as output. The dimension of the vector is the number of the data measured from the object. For example, in order to judge a man, whether he is healthy or unhealthy, we can prepare a vector of parameters such as blood pressure, weight, height, and age and input it into the pattern recognition program, and then, the program gives us a answer "healthy" or "unhealthy" as output.

The pattern is a vector of nth dimension, and the vector space of nth dimension is called pattern space. Patterns derived from the objects which belong to a category occupy a certain subspace in the pattern space. The patterns of another category occupy another area in the pattern space. In the technology of pattern recognition, the boundary between subspaces is decided as first. This is the process to construct the mechanism of pattern recognition. Secondly, after the mechanism is established, we can put unknown pattern into the program and get the answer. The answer is the category to which the unknown pattern belongs.

The boundary between subspaces in a nth pattern super-space is expressed as super-plane of $(n-1)$th dimension. To construct the mechanism of pattern recognition in other words, divide the pattern super-space by super-planes of $(n-1)$th dimension. In many cases, the subspace occupation is complex and it is not easy to set the boundaries. To solve this issue, many kinds of algorithms for multilayered neural networks are used. For the construction of the mechanism, a training method using training patterns is used. By this method, an optimum recognition algorithm is constructed automatically, as if the computer receives the final algorithm by training by itself. RCE network is a kind of neural network, and its feature is to get the mechanism by less training. In this section, an application of RCE network to wood inspection is introduced.

RCE network is basically a three-layered network consisting of input, middle, and output layers. In the training of the RCE network, the mechanism of pattern recognition is established by the following process: (1) A training pattern whose category is already known is placed in the pattern space, and then, a small area of the category is acquired around an input training pattern. (2) By the repeating this process, a body of subspace gathered by the small areas of the category is formed in the pattern space. However, if the area of a category around a training pattern interferes with the area of other categories, then the size of each area is adjusted so that the interference is eliminated. (3) The process is repeated until all the training patterns can be located in the subspace of the right categories. The training process of the RCE network is relatively simple and it requires less training.

Figure 14.80 [47] shows an example of application of the RCE network to the automatic diagnosis of blue-stained wood. By adopting pattern vectors of multiple
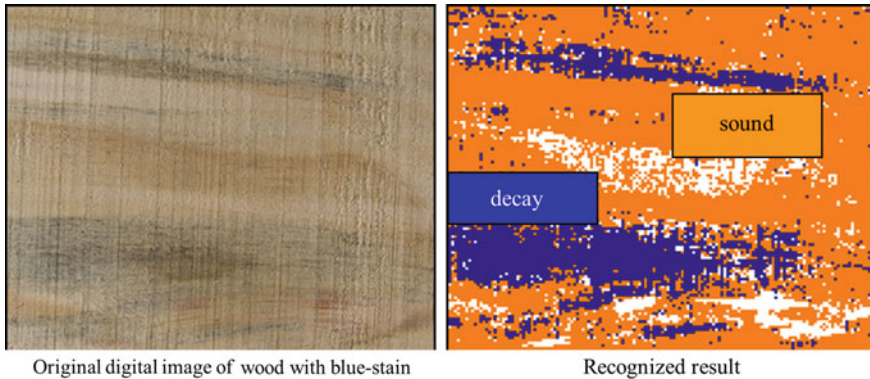
Fig. 14.80 Pattern recognition of 2D image using RCE network. *Note* modified from the literature by Sawada [47]

elements derived from the analysis of color components of pixels, the areas of blue-stained and sound were recognized successfully. By taking the training patterns, mechanisms of diagnosis can be established flexibly according to the local variation and features of the woods. Once the mechanism is established by training, an automatic diagnosis is proceeded quickly by non-touching method and nondestructively.

References

1. Jackson JD (1962) Classical electrodynamics. Wiley, New York
2. Smythe WB (1989) Static and dynamic electricity, 3rd edn. CRC Press, Boca Raton
3. Bell JR, Leonards GA, Dolch WA (1963) Determination of moisture content of hardened concrete by its dielectric properties. In: Proceedings of the American society for testing and material, vol 63
4. Irie H, Yoshida Y, Sakurada Y, Ito T (2008) Non-destructive-testing methods for concrete structures. In: NTT Tech Rev 6(5). https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200805le3.pdf&mode=show_pdf
5. Light ESPAR 2, AIREC Engineering Corporation. http://www.airec.co.jp/products/html/000056.html
6. Taniguchi T, Kobayashi T (2003) An omnidirectional and low-VSWR antenna for the FCC-approved UWB frequency band. In: Proceedings of the 2003 IEEE AP-S international symposium, Columbus, USA, June 2003, pp 460–463
7. Kobayashi T, Takahashi N, Yoshikawa M, Tsunoda K, Tennno N (2007) Measurement of automobile UWB radar cross sections at Ka band. In: Sabath F, Mokole EL, Schenk U, Nittsch D (eds) Ultra-wideband short-pulse electromagnetics 7. Springer, New York, pp 586–592
8. Taniguchi T, Maeda A, Kobayashi T (2007) An omnidirectional and low-VSWR ultra wideband antenna for a frequency band of 6 to 40 GHz. In: Braun CE, Stone AP, Tyo JS (eds) Ultra-wideband, short-pulse electromagnetics 8. Springer, New York, pp 41–48
9. Yang T, Davis WA (2004) Planar half-disk antenna structures for ultra-wideband communications. In: Proceedings of the 2004 IEEE AP-S international symposium, Monterey, USA, June 2004, pp 2508–2511

10. Hirota A, Iwasaki H (2005) Planar trapezoid dipole antenna with ultra wideband characteristics. In: Proceedings of the 2005 IEEE AP-S international symposium, Washington D.C., USA, July 2005, pp 540–543

11. H. Sato, K. Sawaya, N. Arai, Y. Wagatsuma, and K. Mizuno "Broadband FDTD analysis of Fermi antenna with narrow width substrate," in Proceedings of the 2003 IEEE AP-S International Symposium, Columbus, USA, June 2003, pp. 261–264

12. Kim IK, Kidera N, Pinel S, Papapolymerou J, Laskar J, Yook J-G, Tentzeris MM (2006) Linear tapered cavity-backed slot antenna for millimeter-wave LTCC modules. IEEE Antennas Wirel Propag Lett 5:175–178

13. Uehara K, Miyashita K, Natsume K, Hatakeyama K, Mizuno K (1992) Lens-coupled imaging arrays for the millimeter-and submillimete-wave regions. IEEE Trans Microw Theory Tech 40 (5):806–811

14. Mizuno K (2001) Millimeter wave imaging technologies (invited). In: Proceeding of the 2001 Asia-Pacific microwave conference, Taipei, December 2001, pp 394–398

15. Lee RQ, Simons RN (1997) Advances in microstrip and printed antennas, In: Lee HF, Chen W (ed) chapter 9, Wiley, New York

16. Sugawara S, Maita Y, Adachi K, Mori K, Mizuno K (1997) A mm-wave tapered slot antenna with improved radiation pattern. In: IEEE MTT-S international microwave symposium digest, Denver, USA, pp 959–962

17. Sugawara S, Maita Y, Adachi K, Mori K, Mizuno K (1998) Characteristics of a mm-wave tapered slot antenna with corrugated edges. In: IEEE MMT-S international microwave symposium digest, Baltimore, USA, pp 533–536

18. Takagi Y, Sato H, Wagatsuma Y, Sawaya K, Mizuno K (2004) Study of high gain and broadband antipodal fermi antenna with corrugation. International Symposium Antennas and Propagation, vol 1, Sendai, Japan, pp 69–72

19. Wang S, Chen XD, Parini CG (2007) Analysis of ultra wideband antipodal vivaldi antenna design. In: Proceedings of the 2007 Loughborough antennas and propagation conference, Loughborough, UK, April 2007, pp 129–132

20. Gu K, Wang G, Li J (2004) Migration based SAR imaging for ground penetrating radar systems. IEE Proc Radar Sonar and Navig 155:317–325

21. CST STUDIO SUITE 2006B, Computer simulation technology GmbH http://www.cst.com

22. Kaneko S, Oka S, Matsumiya N (2012) Detection of cracks in concrete structures from digital camera images. NTT Technical Review, vol 10, No. 2, 2012. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201202fa3.html

23. Oka S, Mochizuki S, Togo H, Kukutsu N (2009) Inspection of concrete structures using millimeter-wave imaging technology. NTT Tech Rev 7(3). https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200903sf4.html

24. Sheen DM, McMakin DL, Hall TE (2001) Three-dimensional millimeter-wave imaging for concealed weapon detection. IEEE Trans Microw Theory Tech 49(9):1581–1592

25. Ravan M, Amineh RK, Nikolova NK (2010) Near-field microwave holographic imaging: target localization and resolution study. In: 2010 URSI international symposium on electromagnetic theory, Aug 2010, pp 396–399

26. Kak AC, Slaney M (2001) Principles of computerized tomographic imaging. Society of Industrial and Applied Mathematics, Philadelphia

27. Mochizuki S, Oka S, Togo H, Kukutsu N (2010) High-phase-resolution 77-GHz-band radar module for near-field millimeter-wave imaging. In: 2010 IEEE MTT-S international microwave symposium Digest, May 2010, pp 636–639

28. Guidelines for concrete crack inspection, maintenance and reinforcement (2013) Japan Concrete Institute (in Japanese), http://www.jci-net.or.jp/j/publish/book/guide_0080.html

29. Yamaguchi T, Nakamura S, Hashimoto S (2008) An efficient crack detection method using percolation-based image processing. In: Proceeding of the industrial electronics and applications, ICIEA, in Singapore, pp 1875–1880

30. Oka S, Mochizuki S, Togo H, Kukutsu N (2009) A neural network algorithm for detecting invisible concrete surface cracks in near-field millimeter-wave images. In: Proceeding of the

IEEE international conference on systems, man, and cybernetics, neural networks and their applications II, in Texas, pp 3901–3905

31. Ogami M (1987) Hand Book of Image Processing. Tokyo University Publication (in Japanese)
32. Fujita Y, Mitani Y, Hamamoto Y (2006) A method for crack detection on a concrete structure. In: Proceeding of IEEE 18th international conference on pattern recognition, Hong Kong, 20–24 Aug 2006
33. Serre T, Wolf L, Poggio T (2005) Object recognition with features inspired by visual cortex. In: Proceeding of IEEE computer society conference on computer vision and pattern recognition, San Diego, 20–25 Jun 2005
34. Yamaguchi T, Nakamura S, Hashimoto S (2008) An efficient crack detection method using percolation-based image processing. In: Proceeding of industrial electronics and applications, ICIEA, Singapore, pp 1875–1880
35. Markwardt LJ, Wilson TRC (1935) Strength and related properties of woods grown in the United States. In: Technical Bulletin 479, USDA Forest Products Laboratory 99 p
36. Lin T (1967) Review of the electrical properties of wood and cellulose. For Prod J 17(5): 54–60
37. Uemura T (1960) Dielectrical properties of woods as the indicator of the moisture. Bull For For Prod Res Inst 119:95–167
38. Trapp W, Pungs L (1956) Einfluss von Temperatur und Feucht auf das Dielectrische Verhalten von Naturholz im grossen Frequenzbereich. Holzforschung 10(5):144–150
39. Standard of wood processing using high frequency electricity (1960) Technical report of The Institute of Electrical Engineers of Japan, 38
40. Tanaka S, Fujiwara Y, Fujii Y, Okumura S, Togo H, Kukutsu N, Nagatsuma T (2011) Effect of grain direction on transmittance of 100-GHz millimeter wave for hinoki (*Chamaecyparis Obtusa*). J Wood Sci 57(3):189–194
41. Tanaka S, Fujiwara Y, Fujii Y, Okumura S, Togo H, Kukutsu N, Mochizuki S (2013) Dielectric anisotropy of oven- and air-dried wood evaluated using a free space millimeter wave. Journal of Wood Science, Springer, doi:10.1007/s10086-013-1341-7, published April 2013
42. Tanaka S, Fujiwara Y, Fujii Y, Okumura S, Togo H, Kukutsu N, Mochizuki S (2013) Effect of annual rings on transmission of 100 GHz millimeter waves through wood. Journal of Wood Science, Springer. doi: 10.1007/s10086-013-1342-6, published April 2013
43. Fujii Y, Fujiwara Y, Tanaka S, Okumura S, Togo H, Mochizuki S, Kukutsu N (2010) Feasibility of millimeter wave imaging as tool for nondestructive inspection of wood and wooden structures. Proceedings of 35th international conference on infrared, millimeter and terahertz waves, Roma, September 2010. doi:10.1109/ICIMW.2010.5612351
44. Fujiwara Y, Fujii Y, Tanaka S, Okumura S, Togo H, Mochizuki S, Kojima T, Kukutsu N (2012) Feasibility of imaging technology using micro- and millimeter wave for nondestructive inspection of wooden buildings. Proceedings of advanced electromagnetics symposium, AES 2012, proceedings, Paris, April 2012, pp 672–674
45. Fujii Y, Fujiwara Y, Yanase Y, Okumura S, Narahara K, Ngatsuma T, Yoshimura T, Imamura Y (2007) Nondestructive detection of termites using a millimeter-wave imaging technique. For Prod J 57(10):75–79
46. Fujii Y, Fujiwara Y, Yanase Y, Mori T, Yoshimura T, Nakajima M, Tsusumi H, Mori M, Kurisaki H (2013) Development of radar apparatus for scanning of wooden-wall to evaluate inner structure and bio-degradation non-destructively. Adv Mater Res 778:289–294
47. Sawada Y, Matsumoto A, Fujii Y (2015) Pattern recognition of blue stain discoloration appeared on radiata pine boards. Mokuzai Gakkaishi 61(4):274–279

Chapter 15
Principles of Magneto-optic Imaging and Its Applications

Yiming Deng, Yuhua Cheng, Liang Xuan and Zhiwei Zeng

Abstract Magneto-optic imaging (MOI) is an emerging hybrid imaging technique that innovatively generates two-dimensional images by combining electromagnetic and optical energies. It has been widely used in detecting surface and subsurface cracks and corrosion in aircraft skins for nondestructive testing purpose. The merits of the MOI that make it attractive include rapid and large-area inspection, insensitivity to liftoff variation, and easy interpretation in contrast to the complex impedance data of conventional eddy current inspections. The early instrument provides binary images of the anomalies based on eddy current induction techniques and an MO sensor using the Faraday's rotation effect. This chapter will give a comprehensive review on this relatively new sensing technology and discuss the recent advances on MOI including the development of novel optical sources and magnetic film materials for various applications, the integration of MOI sensors with automated systems, and the state-of-the-art post-processing algorithms.

Contents

Y. Deng (✉)
Department of Electrical Engineering, University of Colorado-Denver and Anschutz Medical Campus, 1200 Larimer Street, 80217-3364 Denver, CO, USA
e-mail: yiming.deng@ucdenver.edu

Y. Cheng
University of Electronic Science and Technology, Chengdu, China

L. Xuan
GE Health Care, Chiltern, USA

Z. Zeng
Xiamen University, Xiamen, China

15.1   Introduction

In order to achieve faster imaging speed with higher image resolution, several eddy current-based imaging techniques were developed. One of the most successful methods among them is the magneto-optic imaging (MOI) technique invented by Shih and Fitzpatrick in the year of 1993 [1, 2]. The magneto-optic (MO) imager is widely used in detecting surface and subsurface cracks and corrosion in aircraft skins. The instrument provides binary images of the anomalies based on eddy current induction technique and an MO sensor using the Faraday's rotation effect. Magneto-optic/eddy current imaging (EC-MOI), developed for aging aircraft inspection, is a relatively new NDE technology that produces easy-to-interpret, real-time analog images. Although the technique has gained acceptance in the aircraft maintenance industry for use in detecting surface-breaking and subsurface cracks, emerging safety concerns place increasing demands for improvement in the technology so that it can detect fatigue cracks and/or corrosion when they are about half the length of the current detectable standards. Improving sensitivity of the MOI technology entails improving the sensor performance to achieve greater sensitivity to magnetic fields associated with cracks beneath rivet heads, or for hidden corrosion and subsurface cracks examined at low frequencies. In order to implement any improvements, it is absolutely necessary to have the requisite information on the quantitative values of the leakage magnetic field associated with critical flaws in structures. This requires the development of a numerical simulation model with the capability to produce quantitative values of the leakage flux as well as automated image enhancement and quantification techniques. This chapter will comprehensively summarize the recent advances in sensor development and numerical simulation of the MOI model for a variety of controlled geometrical and experimental parametric variations. The model predictions help in visualizing the field/flaw interaction and thereby optimize the design of the sensor and system.

State-of-the-art data analysis techniques dedicated for MO images are also discussed. The rest of this chapter is organized in this way: Sect. 15.2 describes the general principles underlying the MOI method and system design, as well as a brief history of MOI techniques; Sect. 15.3 addresses the formulation of the governing Maxwell's equations in quasi-static regime for the finite element (FE) analysis of MOI, after which the geometries and simulation results in the presence of corrosion dome and radial cracks are discussed; Sect. 15.4 presents the advanced data analysis techniques developed for MO image characterization; and concluding remarks and future directions are summarized in Sect. 15.5.

15.2   General Principles of Magneto-optic Imager Design

15.2.1   Eddy Current Effect

The physics of eddy current phenomenon is illustrated in Fig. 15.1. The time-varying magnetic flux Φ is generated when the excitation coil is excited by alternating current i. Due to the electromagnetic coupling, eddy current i_2 is produced in the conducting, e.g., metallic specimen that results in a secondary magnetic flux Φ_2 hindering the change of the primary Φ. Since Φ_2 affects the measurable electric parameters of the coil, defects and discontinuities at the surface and subsurface layers and even the properties of materials, e.g., hardness that indirectly determine the secondary flux, can be quantitatively measured by studying the changes of these parameters. More detailed description of eddy current physics can be found in the Electromagnetic Nondestructive Testing Handbook [3] and literature on eddy current testing [4].

15.2.2   Faraday's MO Effect

As shown in Fig. 15.2, when a beam of linearly polarized light passes through a non-optical medium that is optically active in a strong external magnetic field,

Fig. 15.1 Physics of eddy current phenomenon

Excitation Coil

i

χ

Φ

Φ_2

i_2

Specimen
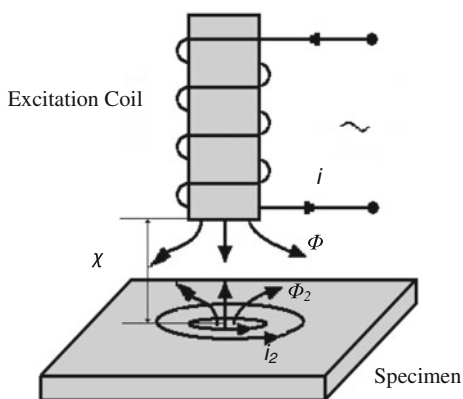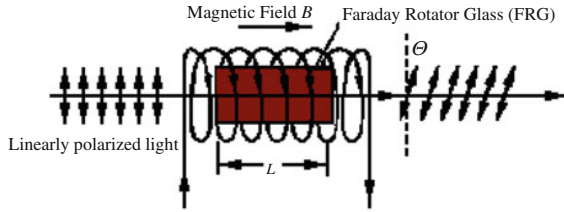
Fig. 15.2 Schematic of the Faraday's rotation effect

whose direction is parallel to the linearly polarized light, the plane of vibration is rotated by θ, according to the Faraday's MO effect. The value of this rotation θ can be expressed as follows:

$$\theta = VBl \tag{15.1}$$

where θ is the angle of Faraday's rotation of light vibration plane in minute arc, B is the magnetic flux density in tesla, and l is the distance the light travels through the medium in meters. And V, called the Verdet constant, is a constant associated with substance and is defined as the rotation angle per unit path per unit field strength.

According to the principle of eddy current testing summarized in Sect. 15.2.1, magnetic field that is either vertical or parallel to the test specimen surface can be induced, whose directions are relied on the excitation coil geometries, when the layered eddy currents flow linearly and distribute uniformly in the examined area of the tested specimen. If there are defects in the specific area, the flow of eddy currents in the defective region will be changed and so will the vertical magnetic field. At the moment, these changes of magnetic field are transformed into relevant variations of light intensity by using MO sensor whose active axis is parallel to the changing magnetic field, and then, the visualization of "real-time" image related to the presence of defects is realized [1, 2, 5]. It is worth noting that the preliminary efforts on imaging and instrumentation prototype were published a few years before the MOI technique became feasible and practical for NDE applications [6–8]. In the past two decades, various MOI systems have been developed for different applications [9–19] and the following section summarizes a brief physics and classification of different MOI systems.

15.2.3 Physics and Classification of MOI Systems

The MOI technique is based on the principles of eddy current effect and Faraday's MO effect, and its systems can be applied to test defects in both ferromagnetic and non-ferromagnetic materials, as illustrated in Fig. 15.3. A foil carrying alternating current works as excitation source, inducing eddy current in the test specimen. If the specimen is intact, the induced currents in the test area flow in the same direction
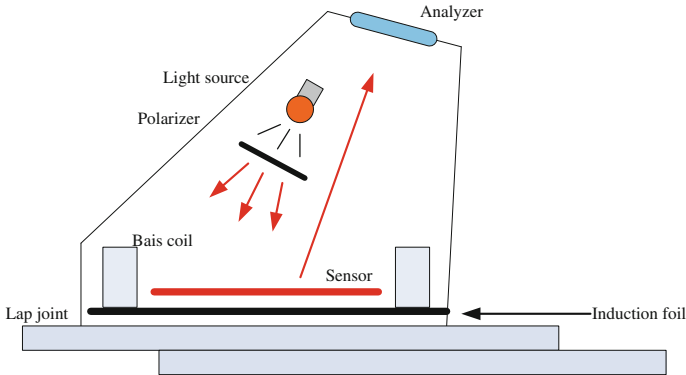
Fig. 15.3 Basic schematic of magneto-optic imaging

parallel to specimen surface and are uniformly distributed. Correspondingly, the magnetic field is parallel to the MO film and the linearly polarized light passing through the MO film twice is not rotated. If there are anomalies in the test area, the eddy current path will be disturbed and its original uniform distribution will be changed; meanwhile, magnetic field that is perpendicular to the surface of test specimen is generated. According to the Faraday's MO effect, when the polarized light is transmitted along the directoin of the "easy axis" of MO film that is parrallel to the eddy current-induced "normal" magnetic field, the polarization plane of the light encounters rotation after passing through the sensor in the area of the anomalies; moreover, the rotation degree lies on the strength of local magnetic field and the specific Faraday's rotation value of the sensor material. Perturbations in magnetic field can be analyzed by measuring the rotation of light polarization. As a result, the MO images can be acquired with pixel intensity correlating with the strength of the normal magnetic fields [20].

In 2006, an MOI technology, called LMOI (linear MO imaging), was patented in Europe by Joubert and Pinassaud as shown in Fig. 15.4 [21]. The LMOI system consists of the optical setup, the eddy current inductor and the linear MO garnet and the CCD camera, as depicted in Fig. 15.4. The optical setup constituted of two polarizers; a LED array is associated with a lens set and designed to produce a uniform and linearly polarized light beam. A specific eddy current inductor is used to excite the inspected material. The presence of a defect in the structure induces variations of the normal component of the field distribution at the surface of the material, which are sensed by a linear MO sensor by Faraday's effect. The sensor is integrated in the optical setup so that the variations of the magnetic field are linearly translated into variations of optical intensities measured by a CCD camera. The advantage of the device is to allow high-speed and high signal-to-noise ratio images to be obtained and furthermore to allow 3D eddy current image inversion to be carried out. Experimental results for a two-layer riveted lap joint have shown that, for both excitation current directions, the surface defects were clearly observable
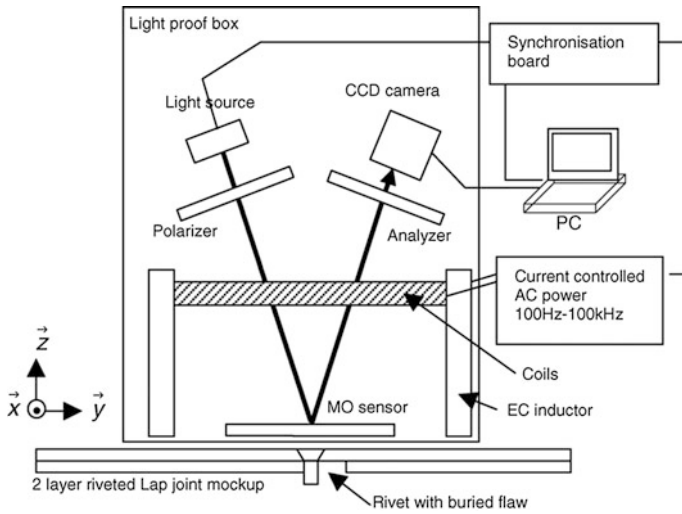
Fig. 15.4 Schematic representation of the LMOI
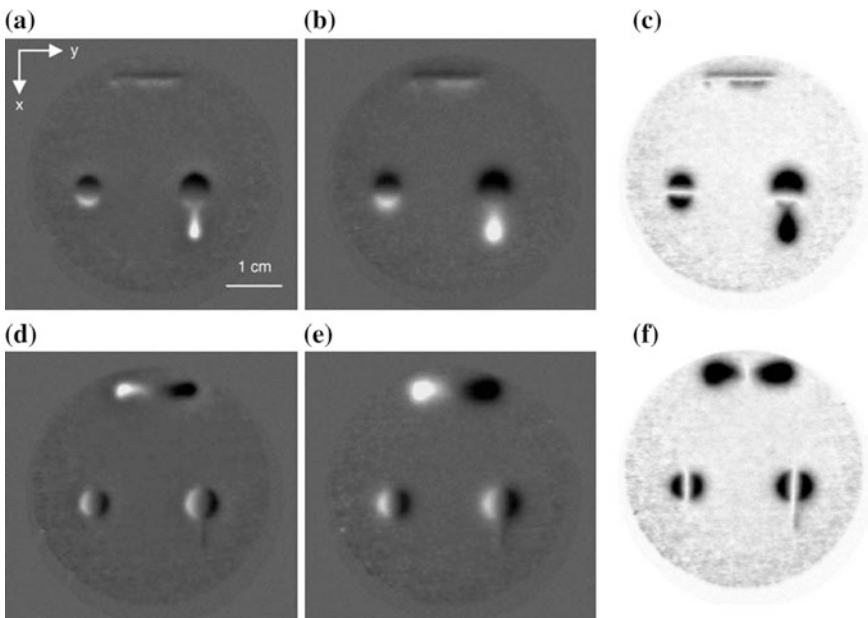
Fig. 15.5 MO/eddy current images of a two-layer riveted lap joint in surface defects (excitation current 5 kHz and 4.5 A). Sound rivet on the left and flawed rivet on the right, **a** and **b** are the real and imaginary parts of the magnetic field (excitation current in the *y*-axis), **c** is the modulus of the image, and **d**–**f** are similar to **a** and **b** except that the excitation current is in the x-axis
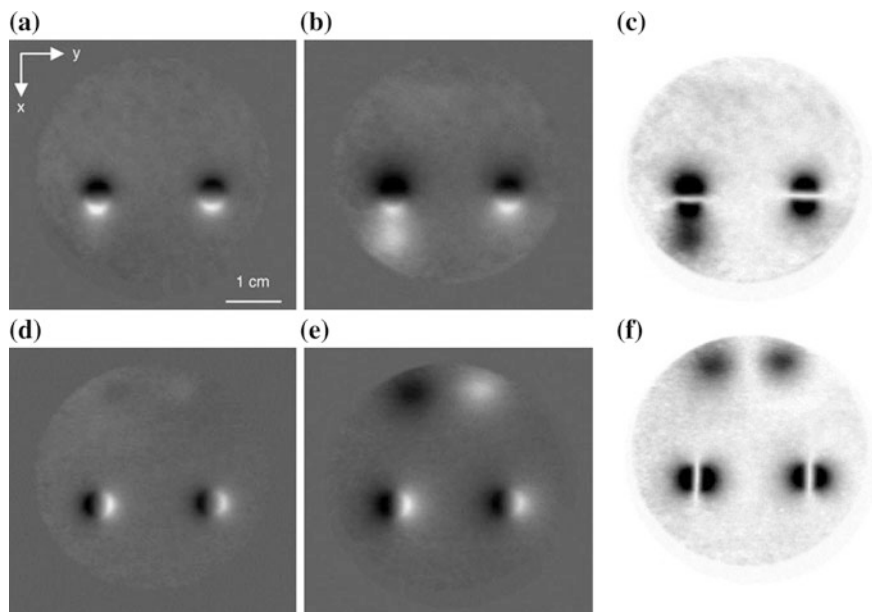
Fig. 15.6 MO/eddy current images of a two-layer riveted lap joint in deep buried defects (excitation current 500 Hz, 8 A). Sound rivet on the right and flawed rivet on the left, **a** and **b** are the real and imaginary parts of the magnetic field (excitation current in the *y*-axis), **c** is the corresponding modulus, and **d**–**f** are similar to **a** and **b** except that the excitation current is in the x-axis

with a high signal-to-noise ratio (no serpentine structures), as depicted in Fig. 15.5. Buried defects were clearly visible for currents perpendicular to the defect orientation, known from Fig. 15.6.

Meanwhile, to expand the range of high-speed real-time MOI, a new MO system was developed by M. Baziljevich in 2012. Following the basic principle of MO imaging, the system setup has made many significant modifications, including the light source, the optical system, the high-speed camera, the frame synchronization, and the custom-made cryostat, as shown in Fig. 15.7. A special source for the external magnetic field has been designed by using a pump solenoid to rapidly excite the field coil. Together with careful modifications of the cryostat, to reduce eddy currents, ramping rates reaching 3000 T/s have been achieved. Using a powerful laser as the light source, a custom-designed optical assembly, and a high-speed digital camera, real-time imaging rates up to 30,000 frames per second.

Attached to the cryostat, the applied field solenoid can ramp from zero to 60 mT during 20 μs, which corresponded to a ramping speed of 3000 T/s. This can be demonstrated in Fig. 15.8. Figure 15.9 reflects detailed flux penetration into the weak area, with the recording speed of 30,000 frames per second. In the picture, the first 8 images (a–h) were sampled at 33-μs intervals and showed a tongue of flux penetrating slowly from the sample edge. The last 4 frames (i–l) show the subsequent flux progression at intervals of 330 μs. It is important to note that the entire series shown
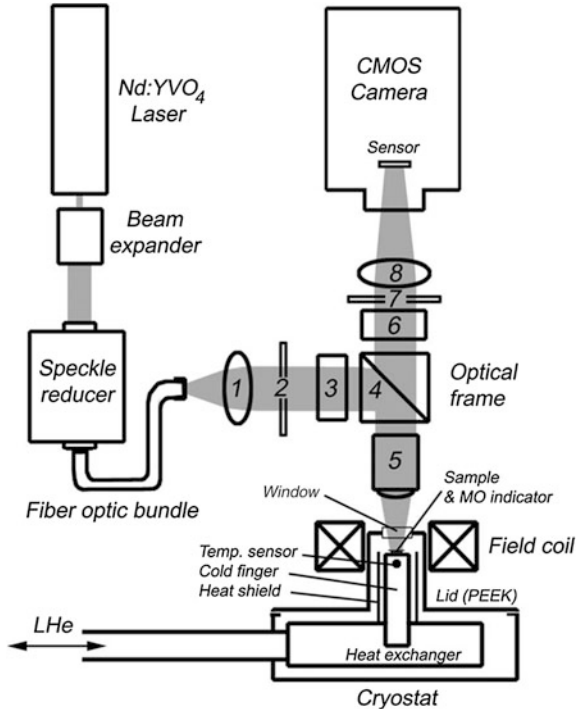
spanned duration of less than 1.5 ms, which was much better than a standard MO setup. After all, test experiments on a BSCCO single crystal have demonstrated MO imaging at 30,000 frames per second, with an acceptable image quality [22].

In 2007, Cheng et al. have achieved an enhanced MOI system by using a laser to improve the sensitivity and image resolution [11]. After that, in 2013, to enhance the MO image quality, they have done some further work including the optimization of the MO sensor, the design of the magnetic excitation device, and the development of the image processing approaches [10]. Figure 15.10 shows the principle configuration of the improved MO testing device, which adopted the semiconductor laser instrument as the light source, and the laser in the system was divided into two beams in horizontal and vertical by the polarization beam splitter. This new MO sensor configuration was designed to reduce the multipath propagation of light and depress the noise in the acquired LMOI effectively. A new form of pulse generator based on Power Factor Correction (PFC) was utilized to provide intermittent excitation current for the coil, which had a lower working temperature of the coil and a higher exciting efficiency than a sinusoidal wave excitation. An innovative wavelet-transform-based image processing approach was also developed, which extracted the contour of the image clearly from the acquired data for better subsurface defect detection.

A specimen under test with an EDM slot was customized for characterizing the newly developed system. The width of the slot was 0.2 mm, and the specimen was scanned from right to left. The MO images are shown in Fig. 15.11a–d. It can be

Fig. 15.8 Oscilloscope traces
showing the fast ramping
profiles of the coil current and
Hall probe voltage at 10 K.
a Using a 20-μs ramp time
and **b** using a 60-μs ramp time
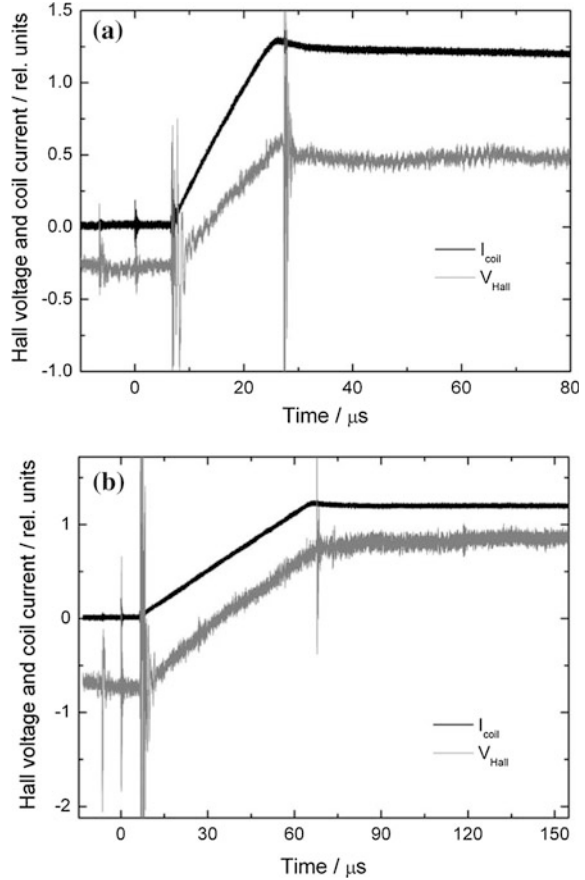
clearly seen that images of the slot gradually appeared from left and then disappeared from right. Figure 15.11c shows the shape of the slot defect very well. This was due to the fact that change of the magnetic field in this position is the largest corresponding to the greatest MO Faraday's rotation angle. The post-processed MO image is shown in Fig. 15.12 using discrete wavelet transform (DWT). Comparing to the conventional MOI systems, the major contribution of this research was to explore the subsurface imaging capability and limits by boosting up the sensitivity and resolution using laser technology [10].

In addition, a high-sensitive scanning laser MOI system has also been developed by Murakami and Tonouchi in 2010. The system was mainly composed of a laser source, galvanometers, and a high-sensitive differential optical detector, as shown in Fig. 15.13. In this system, the laser beam that was linearly polarized by a polarizer was finally focused onto the sample by using an objective lens. And the differential photodetector was composed of a half-wave plate, NPBS, two analyzers, two focusing lenses, two unbiased-type photodiodes, and a differential amplifier. By equipping with the high-sensitive differential photodetector, the developed system
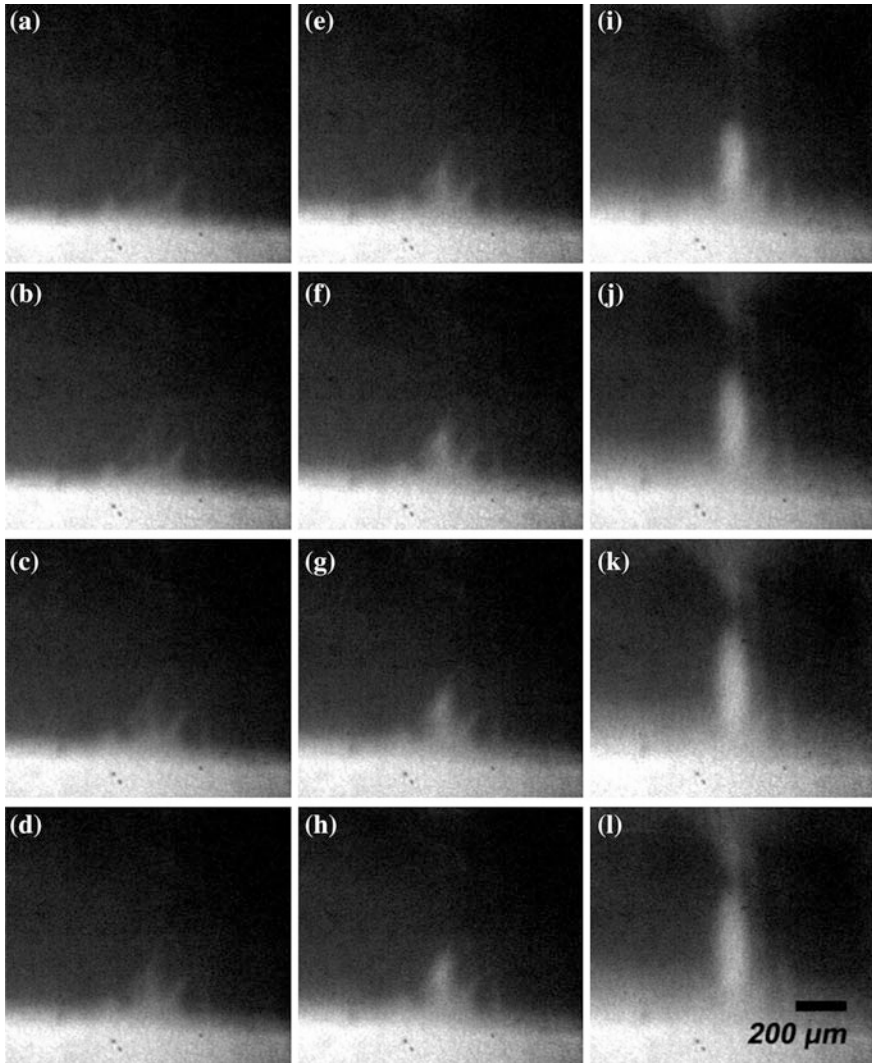
Fig. 15.9 A time series of gradual flux penetration into the weak region at a recording speed of 30,000 fps

can observe the dynamic changes of the local magnetic signal while keeping the sensitivity of about 5 µT without any data processing. Moreover, it can not only realize local signal detection, but also directly evaluate both the strength and direction of a magnetic signal.

To ascertain the actual sensitivity of the developed system, the MO signal of MOD garnet film by applying a square-shaped magnetic field with amplitude of about 50 µT at 100 Hz is observed in Fig. 15.14. It can be seen in the inset that system sensitivity of ∼5 µT can be achieved in the developed system.
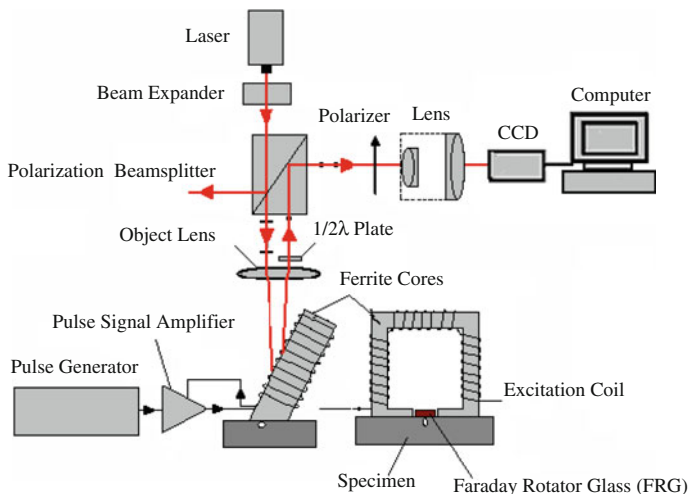
Fig. 15.10 Principle configuration of the laser-based magneto-optic device developed by Cheng and Deng
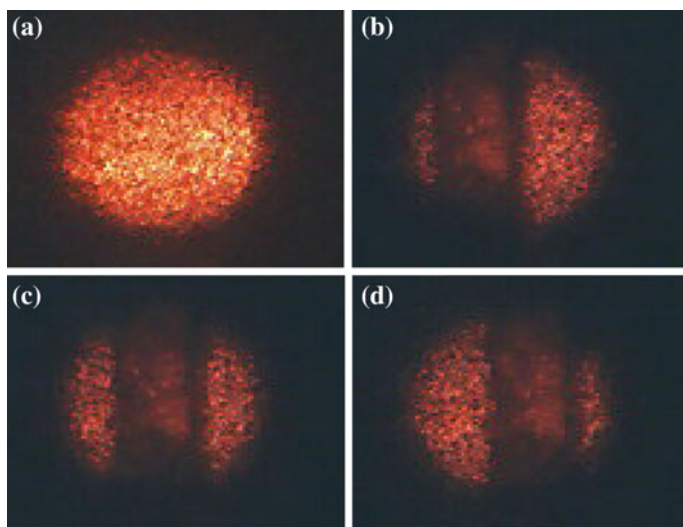
Fig. 15.11 Enhanced LMO images clearly showing the location and shape of subsurface defect

Using this system, MO imaging was carried out on an YBCO ($YBa_2Cu_3O_{7-\delta}$) thin-film strip line of 300 μm wide. At this experimental environment, results in Fig. 15.15a, b have shown the three-dimensional (3D) and two-dimensional (2D) magnetic field distributions around the strip line with the applied dc current of 1 and 3 A, respectively. Figure 15.15c shows the cross-sectional line profiles along the inserted lines of Fig. 15.15a, b. These line profiles have shown a rotational
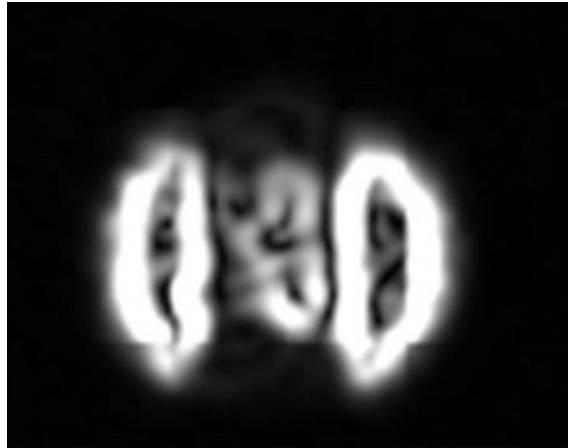
Fig. 15.12 Enhanced MO image after DWT post-processing
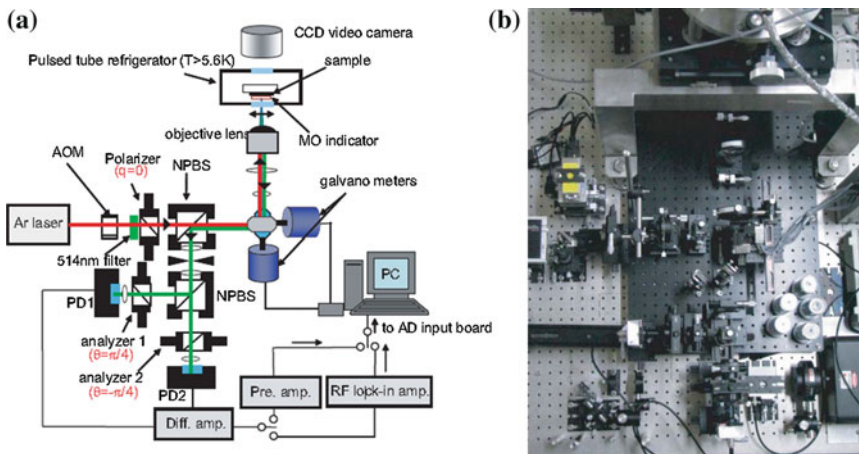
Fig. 15.13 A developed scanning laser MO imaging system with a high-sensitive differential photodetector: **a** optical diagram and **b** photograph of the system

symmetry of the magnetic field distribution around the strip line with almost the same peak intensities in the plus and minus directions. It is noted that the plus and minus peak intensities for $I = 3$ A were correctly three times as large as those for $I = 1$ A, respectively. Therefore, pictures in Fig. 15.15 indicated that the developed MO imaging system enabled us to do quantitative imaging of magnetic field distribution directly, including its direction.

To achieve the local magnetic signal detection in superconductor device, a Josephson vortex flow transistor (JVFT) of YBCO on an MgO bicrystal substrate was prepared. Figure 15.16a shows an optical microscope image of JVFT, where the
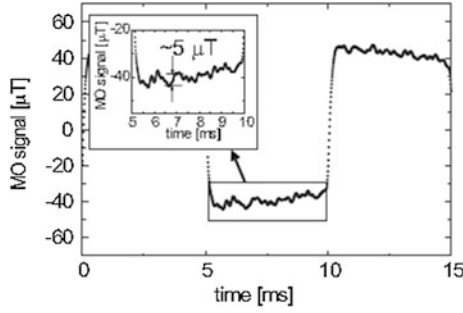
Fig. 15.14 MO signal of the metal organic decomposition garnet film in a perpendicular magnetic field with the amplitude of 46 µT at 100 Hz. Here, the MO signal is reduced into the magnetic field strength
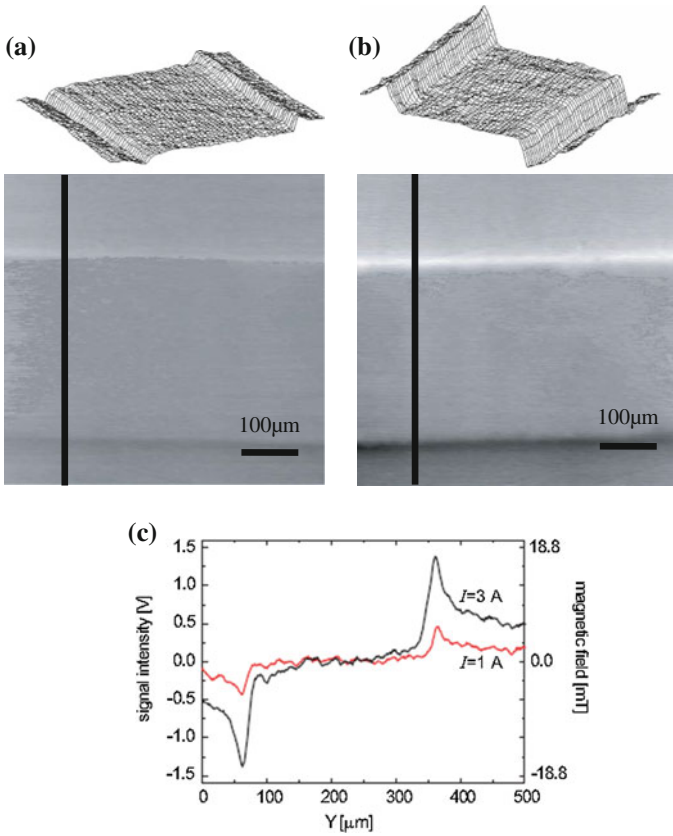
Fig. 15.15 MO images on an YBCO thin-film strip line: **a** and **b** show the 3D and 2D magnetic field distribution around the YBCO strip line with the applied dc current of 1 and 3 A, respectively. **c** The cross-sectional line profiles along the inserted lines of (**a**) and (**b**)
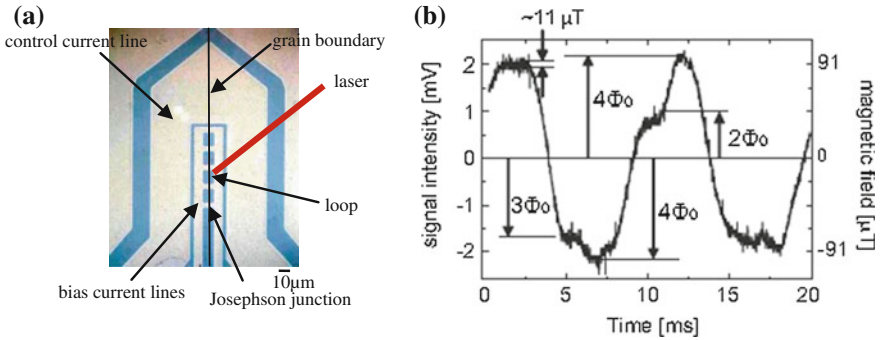
Fig. 15.16 a An optical microscope image of an YBCO JVFT device fabricated on an MgO bicrystal substrate. **b** MO signal from one of the superconducting loops of bias current lines

dotted line shows the position of grain boundary. Figure 15.16b shows the local MO signal in which the laser beam was fixed in one of the superconducting loops. The left axis in the picture has shown the output voltage signal from the differential amplifier, while the right axis has shown the magnetic field strength calculated from the output signal. In the previous system, the noise level was as large as about 50 µT under the same conditions of data processing. However, it was as small as about 11 µT in Fig. 15.16b of this measurement, which indicated that the installed differential photodetector worked very well in reducing the noise level. Furthermore, a clear modulation of MO signal corresponding to the magnetic flux generation inside the loop by applying the control current can be observed, although the irregularity with some step structures was seen. As shown above, it was found that the developed system enables to measure the local magnetic signal directly with high sensitivity and also evaluate the strength and direction of a magnetic signal. As for the sensitivity, if a vibration-free cryostat is used for low-temperature measurements, the sensitivity will be improved up to 5 µT [23]. After two decades when MOI technique was initially developed, the system prototyping and optimization still remains an active research field and some recent work can be found in the literature [12, 18, 24, 25]. For the latest development of MOI magnetic films, the reader can also refer to [26, 27].

15.3 Numerical Modeling of MOI for Aerospace Structure Testing

Meanwhile, to understand the MOI physics, a numerical simulation model that produces quantitative values of the magnetic fields associated with induced eddy currents interacting with structural defects using 3D FE method was developed by Xuan [28] and Zeng et al. [29], which is an essential complement to the instrument development process. The model was applied in simulating aerospace structure testing by the original MO imager. Other modeling efforts on MOI can be referred to the following literature [30–32].

15.3.1 The A,V-A Formulation

A 3D FE model for studying MOI phenomenon has been developed, in which the **A**, **V**-**A** formulation is utilized. The governing equations for time-varying harmonic fields can be derived from the Maxwell's equations and expressed as follows:

$$\nabla \times \frac{1}{\mu} \nabla \times \mathbf{A} + j\omega\sigma\mathbf{A} + \sigma\nabla V = 0 \text{ in } \Omega_1 \tag{15.2}$$

$$\nabla \cdot (j\omega\sigma\mathbf{A} + \sigma\nabla V) = 0 \text{ in } \Omega_1 \tag{15.3}$$

$$\nabla \times \frac{1}{\mu} \nabla \times \mathbf{A} = \mathbf{J}_s \text{ in } \Omega_2 \tag{15.4}$$

In (15.2)–(15.4), Ω_1 and Ω_2 are a partition of the solution domain with Ω_1 and Ω_2 denoting the eddy current region and surrounding free space, respectively, as illustrated in Fig. 15.17. **A** is the magnetic vector potential used in both Ω_1 and Ω_2; V is the electric scalar potential used in Ω_1 only; μ and σ are the permeability and conductivity of the media, respectively.

15.3.2 Numerical Implementation

The solution domain is divided into a number of elements. The potentials at a position inside the eth element are expressed in terms of corresponding functions defined on nodes as follows:

$$\mathbf{A}^e = \sum_{j=1}^{8} N_j^e A_{xj}^e \hat{x} + N_j^e A_{yj}^e \hat{y} + N_j^e A_{zj}^e \hat{z} = \sum_{k=1}^{24} \mathbf{N}_k^e A_k^e \tag{15.5}$$

Fig. 15.17 Region, boundary, and interface in typical eddy current problem
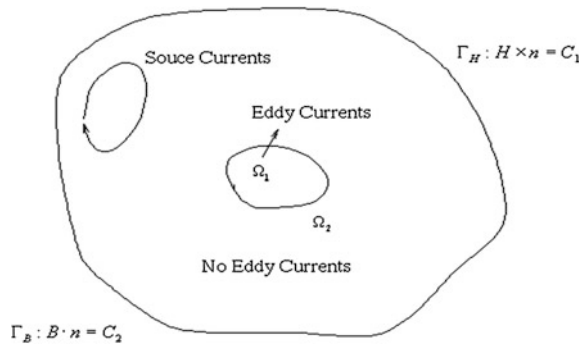
$$V^e = \sum_{j=1}^{8} N_j^e V_j^e \tag{15.6}$$

where

$$\mathbf{N}_k^e = \begin{cases} N_j^e \hat{x} & k = 3j - 2 \\ N_j^e \hat{y} & k = 3j - 1 \\ N_j^e \hat{z} & k = 3j \end{cases} \tag{15.7}$$

and

$$A_k^e = \begin{cases} A_{xj}^e & k = 3j - 2 \\ A_{yj}^e & k = 3j - 1 \\ A_{zj}^e & k = 3j \end{cases} \tag{15.8}$$

where A_{xj}, A_{yj}, A_{zj}, and V_j are the three Cartesian components of the vector potential and the scalar potential at node j, respectively; and N_j is the shape function associated with node j that assumes the value of one at this node and the value of zero at any other node. The vectors $\hat{x}$, $\hat{y}$, $\hat{z}$ are the Cartesian unit vectors.

Using a Galerkin formulation, together with a Coulomb gauge on (15.2), we obtain

$$\sum_{j=1}^{24} \left\{ \int_{\Omega_e} \left(\frac{1}{\mu} \nabla \times \mathbf{N}_i^e \right) \cdot \left(\nabla \times \mathbf{N}_j^e \right) + \left(\frac{1}{\mu} \nabla \cdot \mathbf{N}_i^e \right) \left(\nabla \cdot \mathbf{N}_j^e \right) \mathrm{d}V \right\} A_j^e$$
$$+ \sum_{j=1}^{24} \left\{ \int_{\Omega_e} j\omega\sigma\mathbf{N}_i^e \cdot \mathbf{N}_j^e \mathrm{d}V \right\} A_j^e + \sum_{j=1}^{8} \left\{ \int_{\Omega_e} \sigma\mathbf{N}_i^e \cdot \nabla N_j^e \mathrm{d}V \right\} V_j^e$$
$$- \int_{\partial\Omega_e} \sigma\mathbf{N}_i^e \cdot \left(\frac{1}{\mu} \nabla \times \mathbf{A}^e \times \hat{n} \right) \mathrm{d}S - \int_{\partial\Omega_e} \mathbf{N}_i^e \cdot \left(\hat{n} \frac{1}{\mu} \nabla \cdot \mathbf{A}^e \right) \mathrm{d}S = \int_{\Omega_e} \mathbf{N}_i^e \cdot \mathbf{J}_s^e \mathrm{d}V \tag{15.9a}$$

for $i = 1, 2, \ldots, 24$ and

$$\sum_{j=1}^{24} \left\{ \int_{\Omega_e} j\omega\sigma\nabla N_k^e \cdot \mathbf{N}_j^e \mathrm{d}V \right\} A_j^e + \sum_{j=1}^{8} \left\{ \int_{\Omega_e} \sigma\nabla N_k^e \cdot \nabla N_j^e \mathrm{d}V \right\} V_j^e \tag{15.9b}$$
$$+ \int_{\partial\Omega_e} \sigma N_k^e (-j\omega\mathbf{A} - \nabla V) \cdot \hat{n} \mathrm{d}S = 0$$

$k = 1, 2, \ldots, 8$ where $\hat{n}$ is the outward unit vector, Ω_e is the volume of the element, and $\partial \Omega_e$ is the boundary of Ω_e. Combining these equations, we get a 32×32 elemental matrix equation,

$$[G]^e [U]^e = [Q]^e \tag{15.10}$$

The global matrix is obtained by assembling each elemental matrix $[G]^e$ together. The value at each entry is the sum of values contributed by all the connected elements. Then we have

$$GU = Q \tag{15.11}$$

where G is a complex, symmetric sparse matrix and only nonzero entries need to be stored, U is the unknowns consisting of the electric scalar potential and the three components of the magnetic vector potential at each node, and Q is the load vector incorporating the current source.

15.3.3 Boundary Conditions

In order to get a unique solution, certain boundary conditions are required before solving the system of equations. Either Dirichlet boundary conditions (values of **A** or *V* on the boundary) or Neumann boundary conditions (normal derivative of potentials) should be specified. Since Neumann boundary conditions are usually included implicitly in the finite element formulation, we only need to impose Dirichlet boundary conditions.

Figure 15.18 illustrates the basic geometry used in the chapter which will be more specific in the sections afterward. The boundary conditions for this geometry are listed in Table 15.1. Note that the *x* direction is vertical to the plate surface.

Fig. 15.18 Basic geometry for 3D MOI simulation

19 mm

Current foil

0.35 mm

1.5 mm

Aluminum plate   Corrosion dome
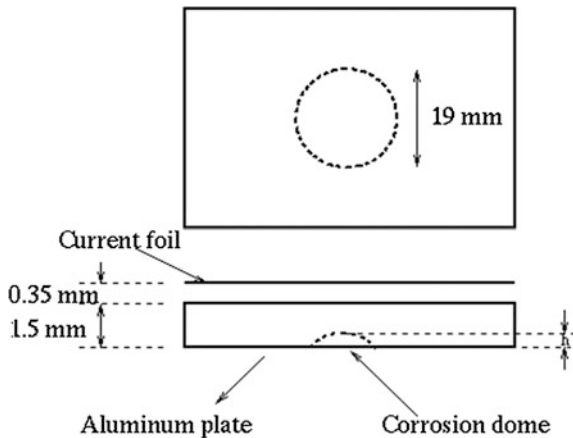
Table 15.1 Boundary conditions

Upper boundary	$A_x = A_y = A_z = 0, V = 0$
Lower boundary	$A_x = A_y = A_z = 0, V = 0$
Left boundary	$A_x = 0, V = 0, A_y, A_z$ free
Right boundary	$A_x = 0, V = 0, A_y, A_z$ free
Front boundary	$A_x = 0, V = 0, A_y, A_z$ free
Rear boundary	$A_x = 0, V = 0, A_y, A_z$ free

15.3.4 Solution of Linear Algebraic Equations

The FE procedure results in a linear algebraic system of equations that must be solved to determine the unknown coefficients of the shape functions. There are two kinds of approaches to solving the system: direct methods and iterative methods. Gaussian elimination is a fundamental method to solve general full or sparse matrix equations. In the FE method, the resulting stiffness matrix is usually sparse and banded, and hence, compared to direct methods, iterative methods are more advantageous with respect to computational cost and data storage. A number of iterative methods have been developed, but few are applicable to general complex-valued matrix equations arising in eddy current problems. Transpose-free quasi-minimal residual (TFQMR) method [33] is one of the applicable approaches. Given a sparse matrix A and vector b, the associated Krylov subspace is $K_n = \text{span}\{b, Ab, A^2b, \ldots, A^{n-1}b\}$. The exact solution $x = A^{-1}b$ could be approximated by $x_n \in K_n$ that minimizes the norm of the residual $r_n = b - Ax_n$. The TFQMR algorithm relaxes this minimization requirement. It aims to minimize a different, data-dependent norm that in practice is similar to $\|r_n\|$. An important issue in iterative methods is its convergence. The convergence of iteration depends on the properties of matrix, such as eigenvalues, singular value, or condition number. Usually, a large condition number will result in large solution error or even non-convergence. Thus, the process of "preconditioning" is essential for the success of iterative methods. In this section, the incomplete LU (ILU) preconditioning technique is used. The overall solution procedure requires $O(N)$ multiplications and additions.

15.3.5 Post-processing

After solving for the potentials, physical and measurable quantities, such as magnetic flux density **B**, electric field intensity **E**, and conduction current density **J**, are calculated using the following formulae:

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{15.12}$$

$$\mathbf{E} = -\mathbf{j}\omega\mathbf{A} - \nabla\mathbf{V} \tag{15.13}$$

$$\mathbf{J} = \sigma\mathbf{E} \tag{15.14}$$

B is calculated and thresholded to generate a binary MO image. J is calculated for the purpose of examining the current distribution in the test specimen. The magnetic flux density **B** can be written in the component form as follows:

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \tag{15.15a}$$

$$B_y = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \tag{15.15b}$$

$$B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \tag{15.15c}$$

In each element (say element e), $\mathbf{B}^e$ is assumed to be constant and could be calculated by

$$B_x^e = \sum_{i=1}^{8} \left(\frac{\partial N_i^e}{\partial y} A_{zi}^e - \frac{\partial N_i^e}{\partial z} A_{yi}^e \right) \tag{15.16}$$

We could get similar form for B_y^e and B_z^e. The current density **J** inside element e is evaluated at the element center. The x component of **J** is as follows:

$$J_x^e = -j\omega\sigma A_x^e - \sigma \frac{\partial V^e}{\partial x} = -j\omega\sigma \sum_{i=1}^{8} N_i^e A_{xi}^e - \sigma \sum_{i=1}^{8} \frac{\partial N_i^e}{\partial x} V_i^e \tag{15.17}$$

The components J_y^e and J_z^e can be calculated similarly.

15.3.6 Complex Form of Rotating Current

For a linear sinusoidal current excitation with angular frequency of ω, the transient form of current is as follows:

$$j_0(y, z, t) = \text{Re}\left[J_0(y, z, \omega)e^{j\omega t} \right] = J_0(y, z, \omega)\cos(\omega t) \tag{15.18}$$

For a rotating current excitation, it is the summation of two linear currents that are orthogonal to each other both in direction and in phase. In time domain, it can be expressed as follows:

$$\begin{aligned} \bar{J}(y, z, t) &= J_0(y, z, \omega) \cos(\omega t)\, \hat{y} + J_0(y, z, \omega) \sin(\omega t)\, \hat{z} \\ &= \mathrm{Re}\big\{ [J_0(y, z, \omega)\hat{y} - jJ_0(y, z, \omega)\hat{z}] e^{j\omega t} \big\} \end{aligned} \tag{15.19}$$

So the corresponding complex form in frequency domain should be

$$\bar{J}(y, z, \omega) = J_0(y, z, \omega)\hat{y} - jJ_0(y, z, \omega)\hat{z} \tag{15.20}$$

15.3.7 Simulation Geometry: Corrosion Dome in Test Sample

The inspected sample consists of one or two aluminum plate layers each of thickness 1.5 mm with a corrosion dome of diameter 19 mm and height h introduced in one of the aluminum plates. Figure 15.19a shows the geometry with the corrosion dome at the bottom of a single layer. Figure 15.19b shows the geometry with the corrosion dome at the bottom of the 1st (upper) layer. Figure 15.19c shows the geometry with the corrosion dome at the top of the 2nd (lower) layer. Figure 15.20 shows the mesh corresponding to Fig. 15.19a. A variety of simulations were conducted using author-optimized meshes to study the effect of location of corrosion; excitation frequency; corrosion depth; air gap between layers; slope of corrosion dome; and edge effect.

15.3.7.1 Effect of Layers and Corrosion Depth

Figure 15.21a–c shows the normal component of the magnetic flux densities in the cases of (Fig. 15.21a) corrosion dome at the bottom of a single layer, (Fig. 15.21b) corrosion dome at the bottom of the 1st layer, and (Fig. 15.21c) corrosion dome at the top of the 2nd layer. In each case, the height of corrosion dome is 20 % of the thickness of an aluminum layer. We find that the magnetic flux density in the case of a single layer is more diffused than that obtained in the cases of double layers. The difference between the cases of the corrosion dome at the bottom of the 1st layer and corrosion dome at the top of the 2nd layer is not very significant. The presence of the second layer reduces the strength of the induced eddy current in the first layer, resulting in the decrease of magnitude of the vertical magnetic flux density B_x. Figure 15.22 shows the peak values of vertical magnetic flux density as a function of depth (h) of the corrosion dome. The results agree with the

Fig. 15.19 Geometries of test samples with corrosion dome: **a** corrosion dome at the bottom of a single layer, **b** corrosion dome at the bottom of the 1st layer, and **c** corrosion dome at the top of the 2nd layer
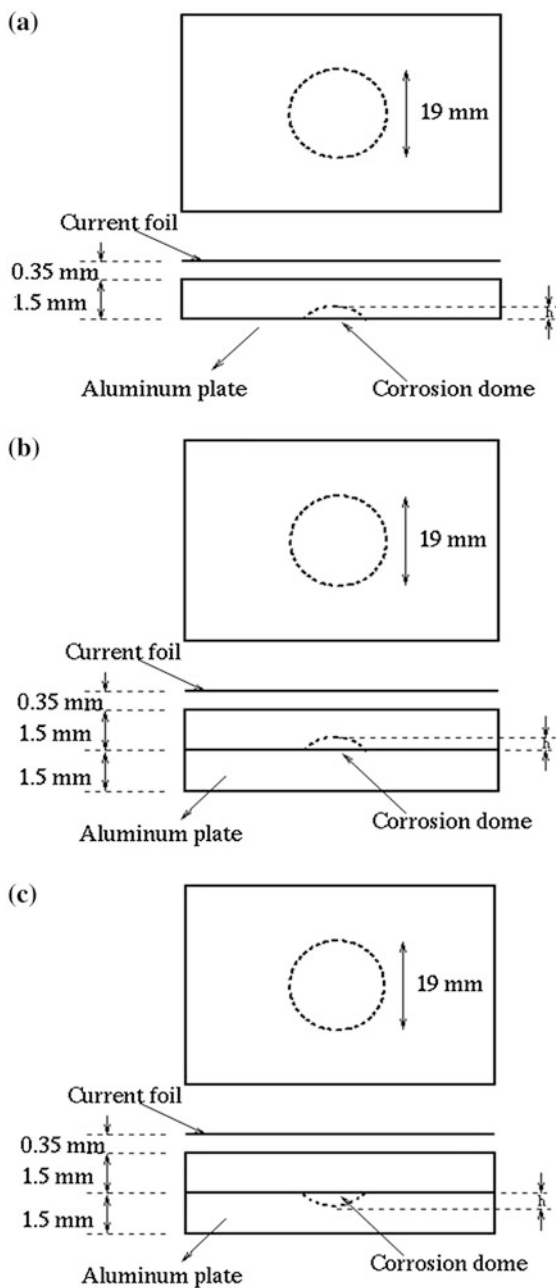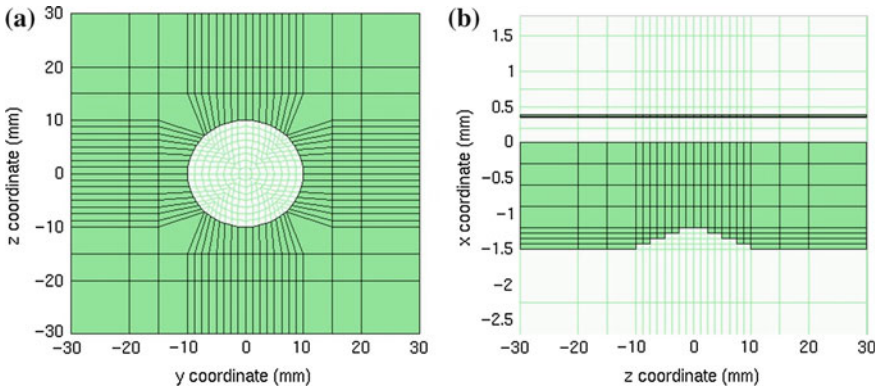
Fig. 15.20 Mesh with corrosion dome at the bottom of a single layer: **a** top view and **b** side view
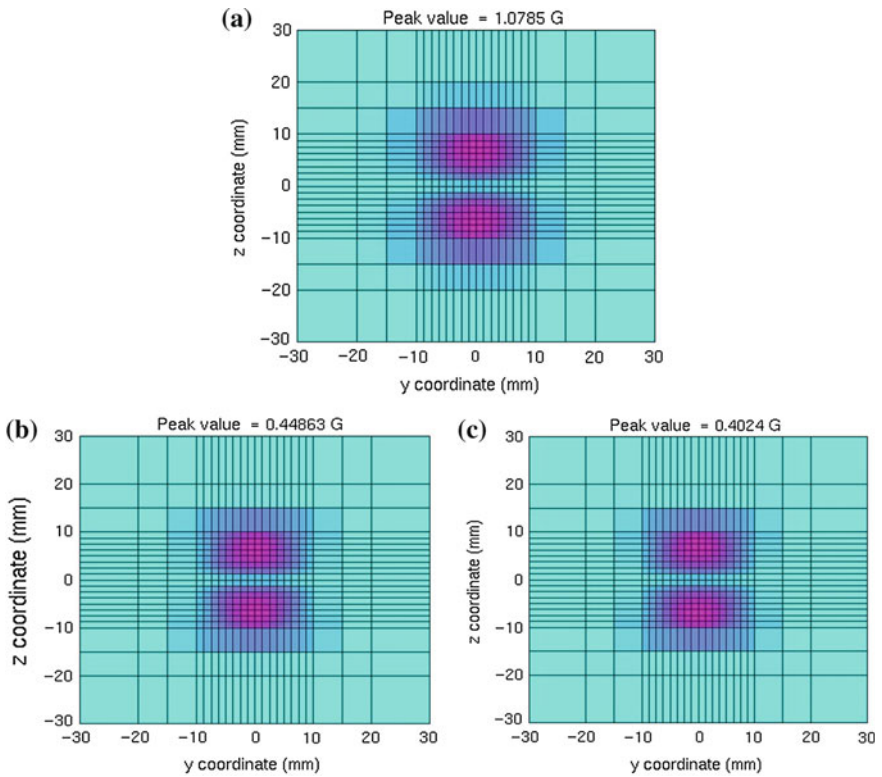
Fig. 15.21 Vertical magnetic flux densities. The height of corrosion dome is 20 % of an aluminum plate, and the frequency is 1 kHz: **a** corrosion dome at the bottom of a single layer; **b** corrosion dome at the bottom of the 1st layer; and **c** corrosion dome at the top of the 2nd layer

Fig. 15.22 Peak value of vertical magnetic flux density versus depth of corrosion dome and frequency at 1 kHz. *Corrosion 1* corrosion dome at the bottom of a single layer; *Corrosion 2* corrosion dome at the bottom of the 1st layer; and *Corrosion 3* corrosion dome at the top of the 2nd layer
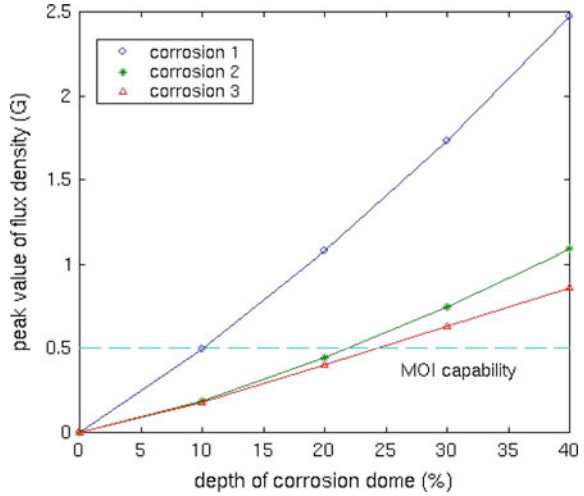
Fig. 15.23 Peak values of flux leakage versus excitation frequency. *Corrosion 1* corrosion dome at the bottom of a single layer; *Corrosion 2* corrosion dome at the bottom of the 1st layer; and *Corrosion 3* corrosion dome at the top of the 2nd layer
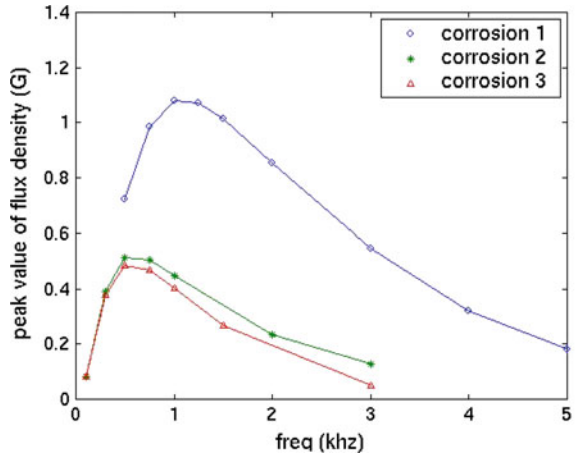
experimental observation that the MOI instrument can detect a 10 % corrosion dome in a single-layer plate, but can only detect a 20 % corrosion dome in the presence of second-layer plate.

15.3.7.2 Effect of Frequency

Figure 15.23 shows the peak value of flux leakage as a function of excitation frequency ranging from 0.1 to 5 kHz, with the depth of corrosion dome 20 % of the aluminum plate. As we can see, the optimal frequency is 1 kHz for a single-layer case and is reduced to 0.5 kHz when a second layer is present.

Fig. 15.24 Peak values of vertical magnetic flux density versus air gap width, where *Corrosion 1* corrosion dome at the bottom of the 1st layer and *Corrosion 2* corrosion dome at the top of the 2nd layer
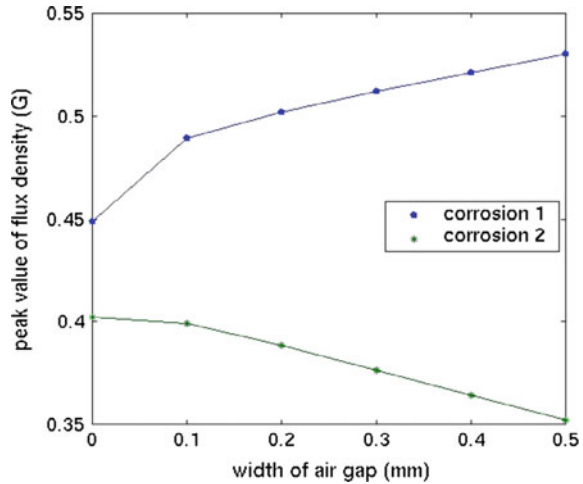
15.3.7.3 Effect of Air Gap Between Two Layers

Typically, an air gap exists between different layers. Figure 15.24 shows the relation of signal peak value with respect to the air gap width ranging from 0.1 to 0.5 mm. The depth of corrosion dome is 20 % of the aluminum plate thickness. With increasing air gap width, the peak magnitude of vertical magnetic flux density increases if the corrosion dome is at the bottom of the 1st layer, since this situation moves toward the single-layer case. When the corrosion dome is located at the top of the 2nd layer, the corresponding peak value of magnetic flux density decreases with air gap width since increasing the width of the air gap shifts the corrosion dome away from the excitation source. Both curves are almost linear in the region of [0, 0.5] mm except for the point where the air gap width is 0. This behavior is due to the fact that the presence of an air gap (even the smallest air gap) materially affects the current distribution in the plates. So the peak value jumps up when the air gap width is increased from 0 to a small value, regardless of the location of the corrosion dome.

15.3.7.4 Effect of Corrosion Slope

The side views of corrosion domes with different slopes are shown in Fig. 15.25a–d. All corrosion domes have the same depth that is 20 % of an aluminum plate. Figure 15.26a–d shows the corresponding normal components of magnetic flux densities. We can see that the MO images of flux leakage vary widely with respect to different slopes of the corrosion dome. The smaller the slope is, the more blurred the images are. The images can therefore be used to characterize the shape of corrosion damage.
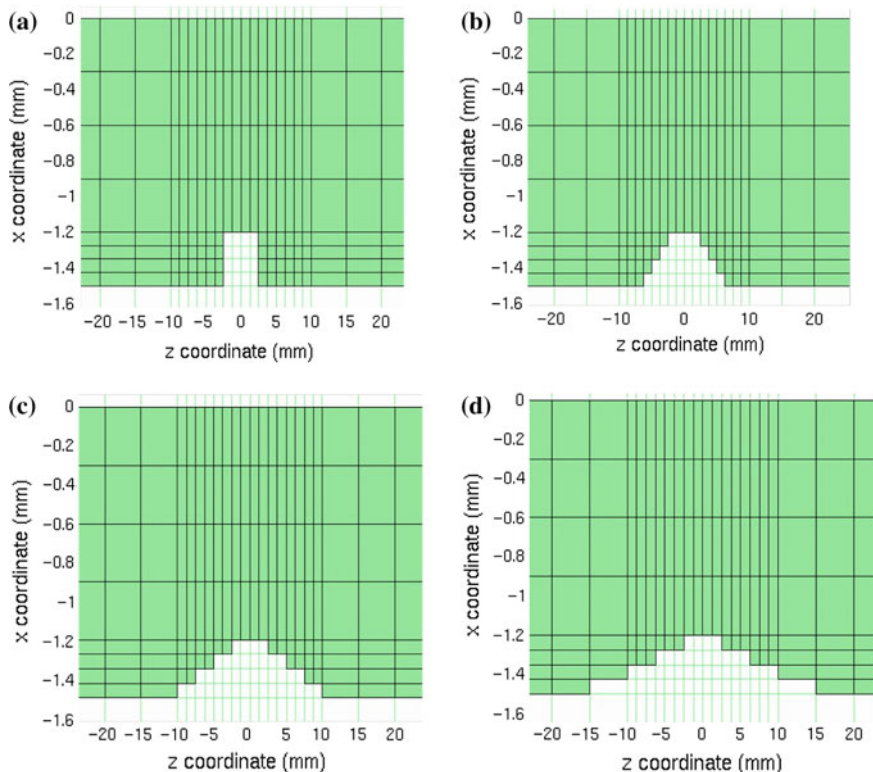
Fig. 15.25 Side views of geometries and meshes of corrosion domes with different slopes

15.3.7.5 Effect of Edge Close to Corrosion

In practice, corrosions may occur near a lap joint, where an edge discontinuity lies in the vicinity of corrosion in the second layer of the sample. A 20 % corrosion dome at the bottom of the 1st layer is modeled in the vicinity of a z-edge or y-edge in the 2nd layer. If the excitation current is along the y direction, then z-edge is perpendicular to it and y-edge is parallel to it. Figure 15.27 shows the geometry with a z-edge (edge is perpendicular to induction current source). Side views of the corrosion domes and edges along with the corresponding images of the normal components of the magnetic flux densities are shown in Figs. 15.28 and 15.29. The cross-sectional views shown are y-x cross section for the z-edge case and z-x cross section for the y-edge case. The edge is located at a distance of 5 or 10 mm to the corrosion center, respectively.

We can see that the MO image of corrosion dome is shifted away from the center when a z-edge is present, while the y-edge is seen in the MO image. It is also observed that in the presence of edges, the peak value of vertical flux density is enlarged, especially in the case of z-edge. We can express the peak value of vertical

(a) Peak value = 0.70485 G

(b) Peak value = 1.0406 G

(c) Peak value = 1.0785 G

(d) Peak value = 0.98327 G
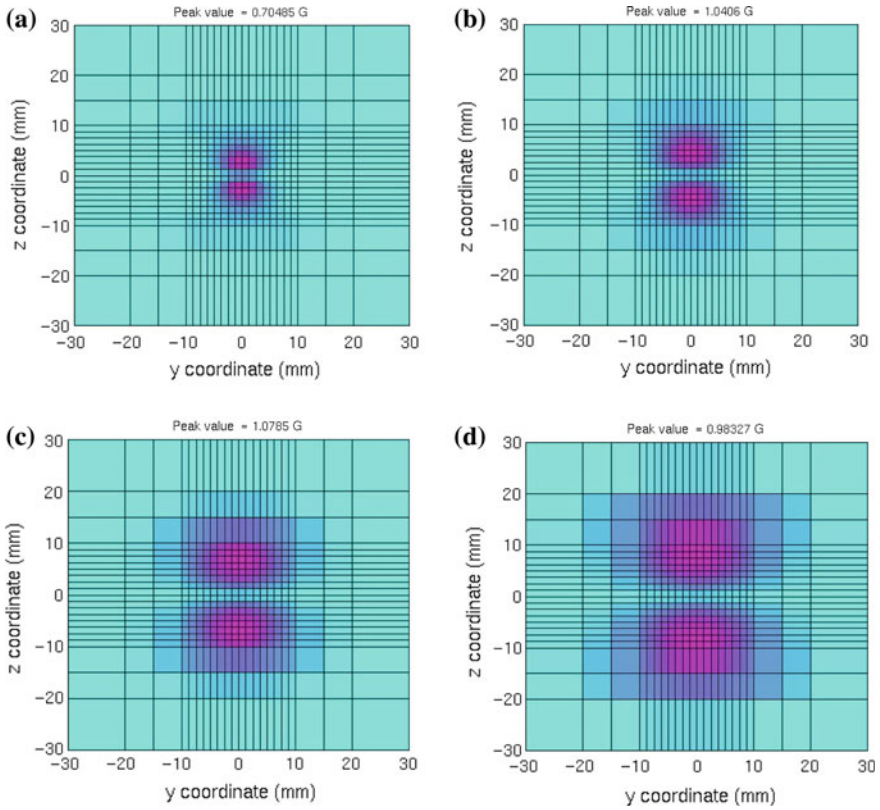
Fig. 15.26 Vertical flux densities corresponding to geometries shown in Fig. 15.25a–d

Fig. 15.27 Geometry of double-layer aluminum plates with z-edge

19 mm

current foil

0.35 mm

1.5 mm

1.5 mm

aluminum plate

z-edge

corrosion dome

h
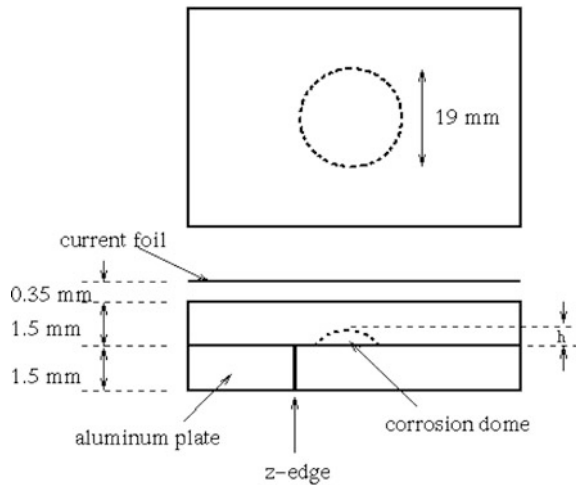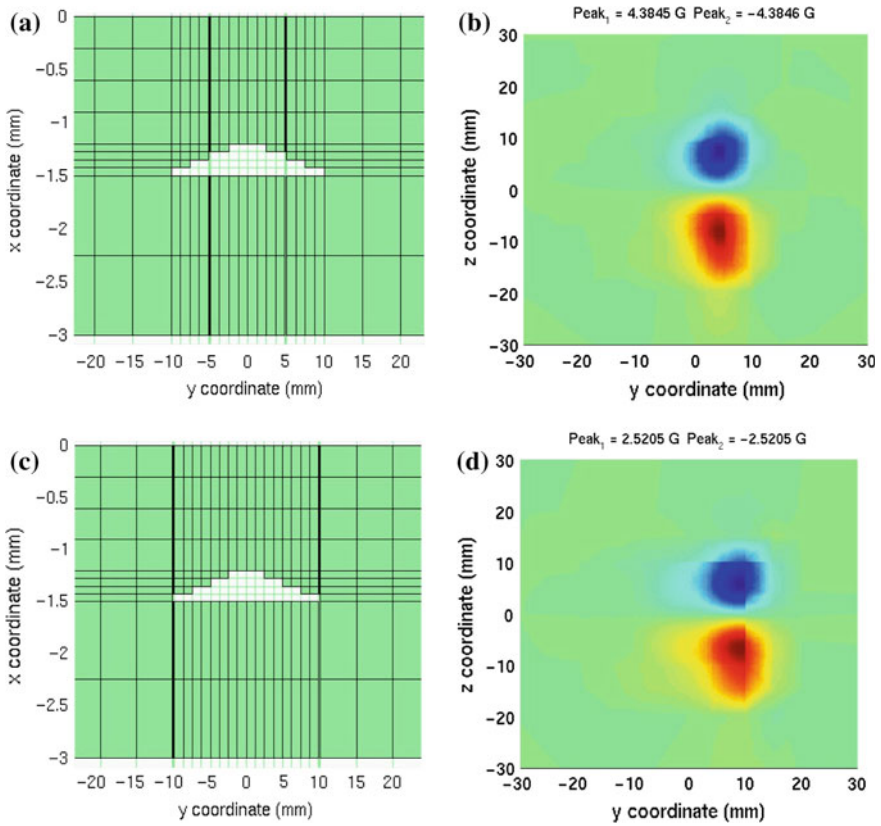
Fig. 15.28 Geometries with z-edges and the associated vertical magnetic flux densities: **a** geometry with z-edge 5 mm from the corrosion center, **b** vertical magnetic flux density corresponding to (**a**), the positive and negative peak values being 4.3845 G and −4.3846 G, respectively, **c** geometry with z-edge 10 mm from the corrosion center, and **d** vertical magnetic flux density corresponding to (**c**), the positive and negative peak values being 2.5205 G and −2.5205 G, respectively

flux density as a function of the distance (h) between edge and corrosion center, as shown in Fig. 15.30. The peak value of vertical flux density decreases with h for the z-edge case, but this relationship becomes complex in the case of a y-edge. A z-edge normal to excitation current direction affects the peak value more distinctly than a y-edge which is along the excitation current direction.

15.3.8 Simulation Geometry: Rivet with Radial Crack

Figure 15.31 shows a vertical section of the geometry under investigation in this section. A buried crack in the second layer of the aluminum plates is considered as
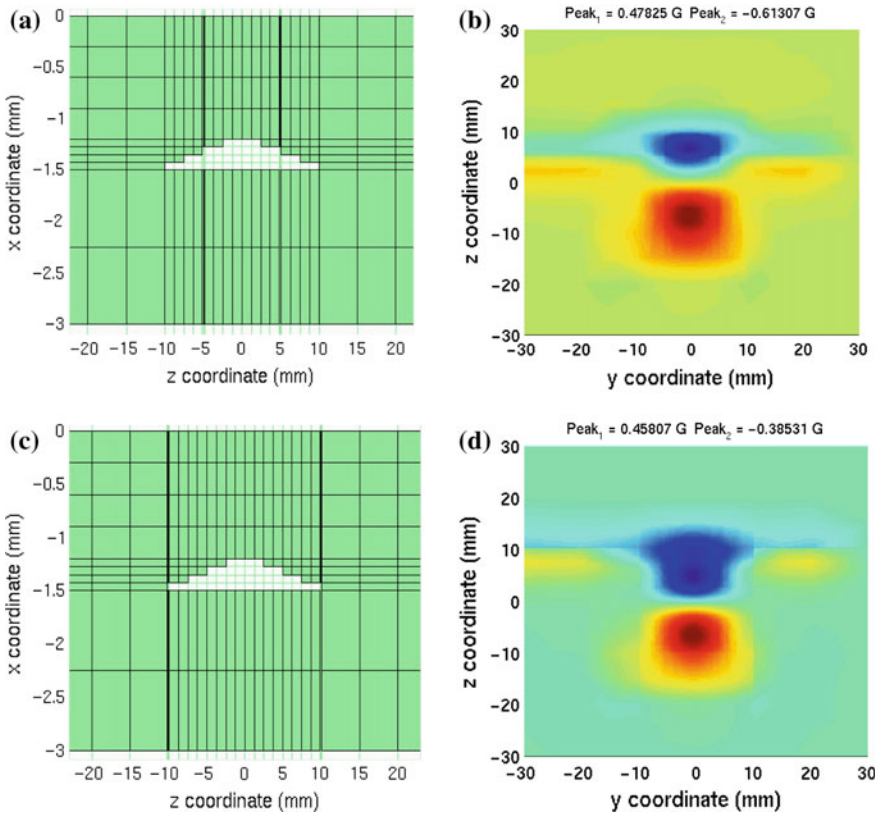
Fig. 15.29 Geometries with *y*-edges and the associated vertical magnetic flux densities: **a** geometry with *y*-edge 5 mm from the corrosion center, **b** vertical magnetic flux density corresponding to (**a**), the positive and negative peak values being 0.4783 G and −0.6131 G, respectively, **c** geometry with *y*-edge 10 mm from the corrosion center, and **d** vertical magnetic flux density corresponding to (**c**), the positive and negative peak values being 0.4581 G and −0.3853 G, respectively

the critical flaw to be detected. The crack shown in the figure is of rectangular shape. The last subsection in this section describes the results of testing a crack that is of wedge shape. The rectangular crack has a height of 1 mm and width of 0.1 mm. The radial length of the crack is varied from 3 to 8 mm for investigating parametric effects. Figure 15.32 is the mesh corresponding to the geometry in Fig. 15.31. A variety of simulations were conducted.

15.3.8.1 Model Validation

Comparison of experimental and numerical MO images in Fig. 15.33a, b validates the FE model.
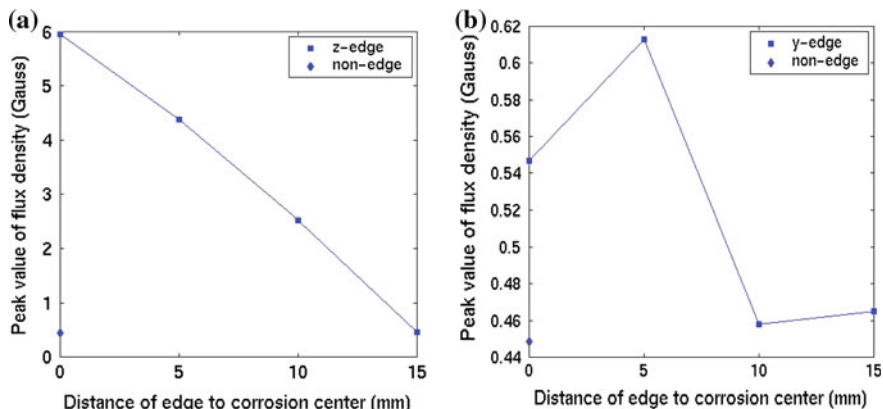
Fig. 15.30 Peak value versus distance between edge and corrosion center: **a** *z*-edge and **b** *y*-edge
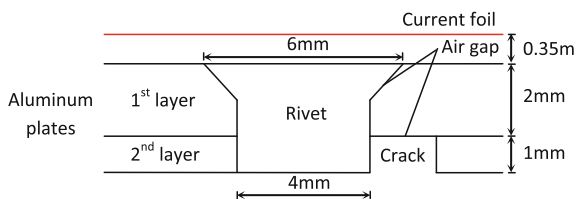
Fig. 15.31 Geometry of rivet with radial crack
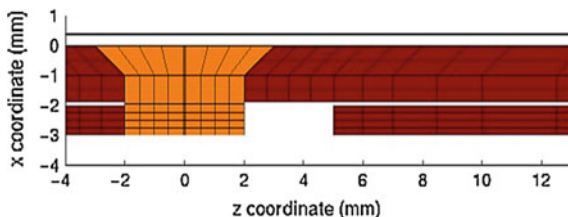
Fig. 15.32 Mesh with radial crack under fasterner

15.3.8.2 Effect of Finite Foil

Figure 15.34 shows the normal component of the predicted magnetic flux densities with the edge effect due to the finite size of the foil. The effects of finite foil versus infinite foil are summarized as follows: (1) The signal peak value is much smaller in case of using finite foil compared with that using infinite foil; (2) the binary images with infinite foil and those with finite foil are comparable if the threshold is chosen to be a percentage of the peak value instead of an absolute value (see Fig. 15.35a, b); and (3) increasing frequency reduces the edge effect (see Figs. 15.35b and 15.36b).
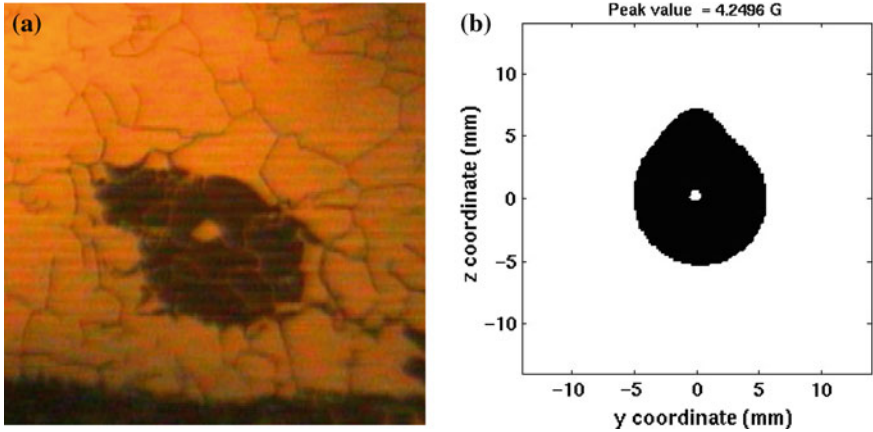
Fig. 15.33 MO image for rivet head with a 5-mm radial crack under rotating current excitation at a frequency of 3 kHz: **a** experimental and **b** numerical results

Fig. 15.34 Field distribution with finite foil with a frequency of 1.5 kHz; crack length is 5 mm
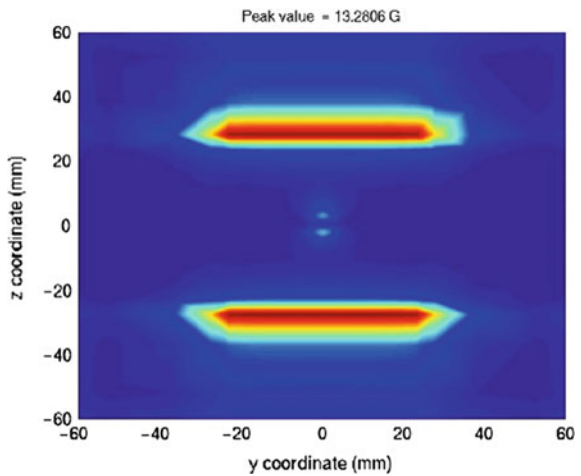
These observations imply that the field distributions for finite and infinite foils are similar in shape. The effect of the foil edge is to produce a field opposite to that at the rivet site, thus reducing the magnitude.

15.3.8.3 Effect of Tear Strap

Figure 15.37a, b illustrates how a tear strap is attached to the aluminum plates in either y or z direction. The tear strap geometry was modeled, and the resulting normal flux density distribution was predicted. The continuously valued flux density was thresholded by an appropriate bias coil field, resulting in a binary image
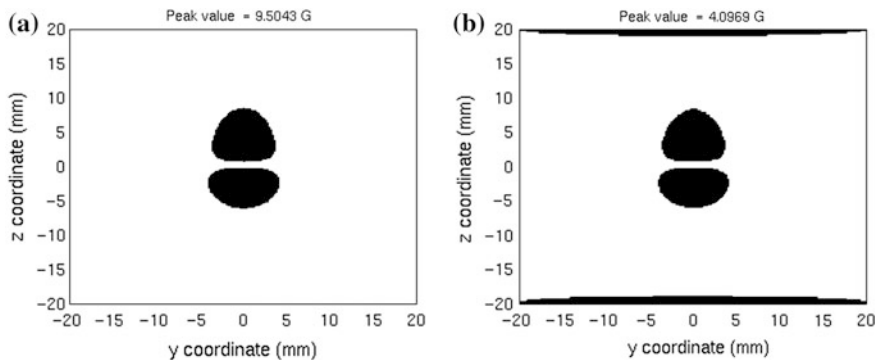
Fig. 15.35 Binary images in the region of interest with a frequency of 1.5 kHz and threshold of 25 % of the corresponding peak value; crack length is 5 mm: **a** with infinite foil, peak value of vertical magnetic flux density being 9.5043 G and **b** with finite foil, peak value of vertical magnetic flux density being 4.0969 G
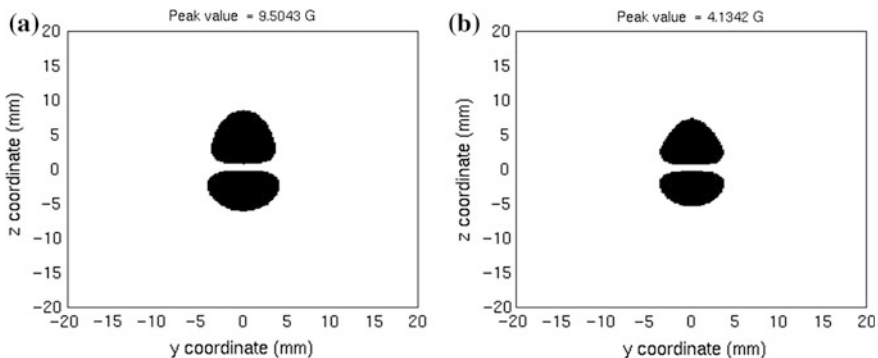
Fig. 15.36 Binary images in the region of interest with a frequency of 3 kHz and threshold of 25 % of the corresponding peak value; crack length is 5 mm: **a** with infinite foil, peak value of vertical magnetic flux density being 9.5043 G, and **b** with finite foil, peak value of vertical magnetic flux density being 4.1342 G
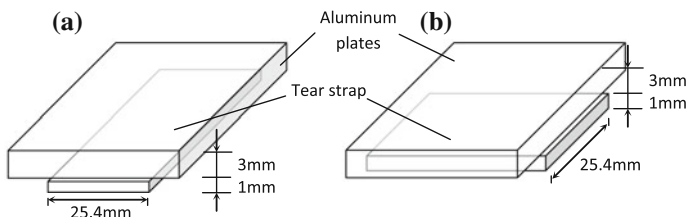
Fig. 15.37 Tear straps under aluminum plates: **a** tear strap in z direction (perpendicular to the crack orientation) and **b** tear strap in y direction (parallel to the crack orientation)
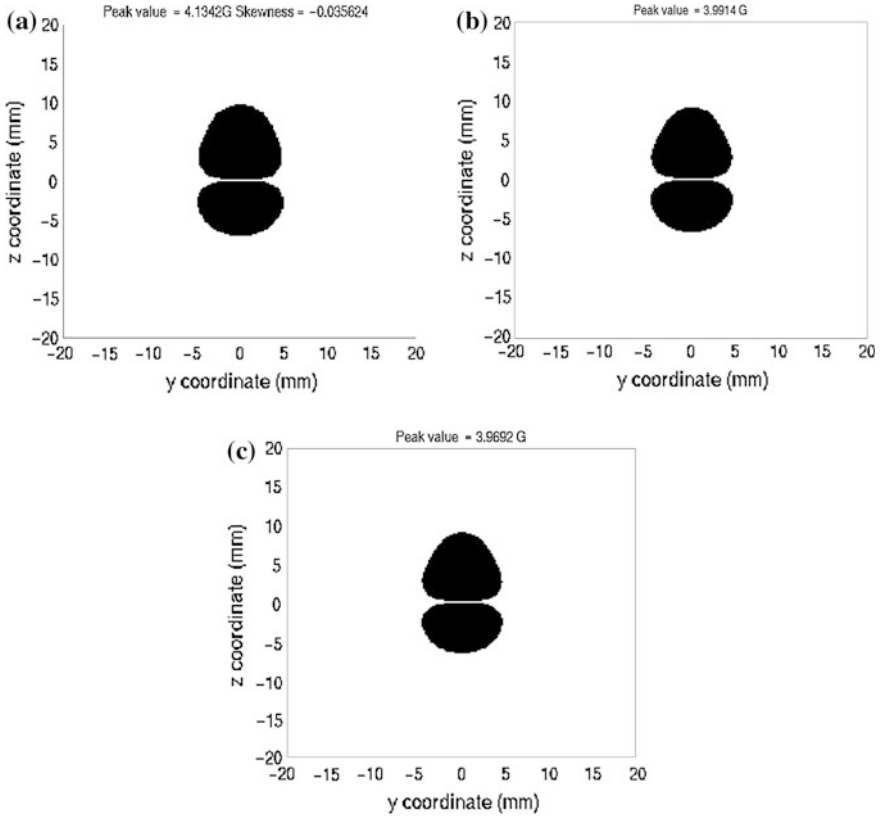
Fig. 15.38 Binary images in the region of interest, with crack length of 5 mm, frequency at 3 kHz, finite foil, and threshold of 0.5 G: **a** no tear strap, peak value of normal magnetic flux density being 4.1342 G, **b** tear strap in z direction, peak value of normal magnetic flux density being 3.9914 G, and **c** tear strap in y direction, peak value of normal magnetic flux density being 3.9692 G

similar to that observed in an experimental system. As shown in Fig. 15.38, tear strap does not have a significant effect on resulting image pattern. However, a tear strap decreases the signal peak value, which is believed to be the principal effect or the tear strap. The corresponding peak values in Figs. 15.38a–c are 4.13, 3.99, and 3.97 G, respectively. The reduction of peak value by the tear strap is not significant because the skin effect reduces the influence of tear strap which is beyond the skin depth of aluminum at a frequency of 3 kHz.

15.3.8.4 Effect of Rotating Versus Linear Current

A linear excitation current results in two lobes along the direction orthogonal to the current direction. In the case of rotating excitation current, the result is equal in all orientations, and hence, this results in a circular ring-shaped image. In the case of
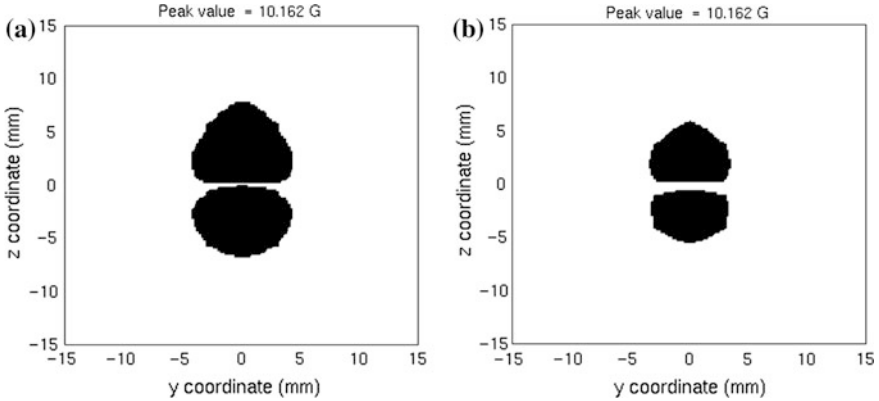
Fig. 15.39 Binary images in the region of interest with linear excitation current perpendicular to crack at a frequency of 5 kHz; crack length is 5 mm: **a** threshold of 1 G and **b** threshold of 2 G, peak value of normal magnetic flux density being 10.162 G
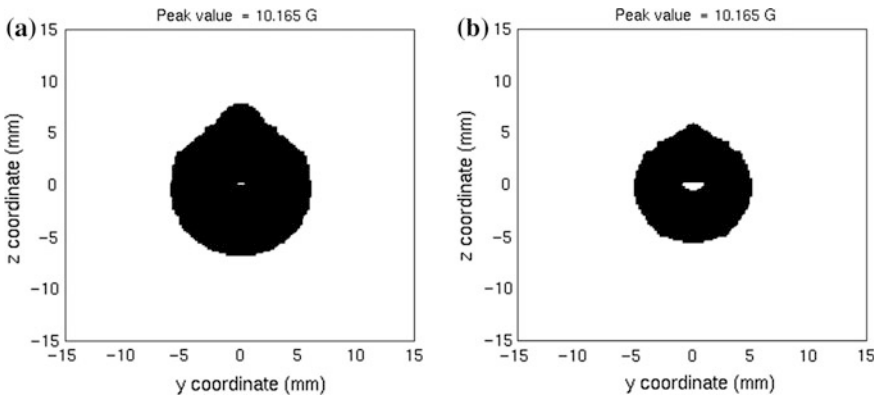
Fig. 15.40 Binary images in the region of interest with rotating excitation current at a frequency of 5 kHz; crack length is 5 mm: **a** threshold of 1 G and **b** threshold of 2 G, peak value of normal magnetic flux density being 10.165 G

linear excitation current, cracks that are parallel to the current direction are not detected and cracks that are perpendicular to the current result in maximum signal (Fig. 15.39). In contrast, rotating current excitation can detect cracks at all angles. There is not a significant difference in signal peak value between linear and rotating excitations (Fig. 15.40).
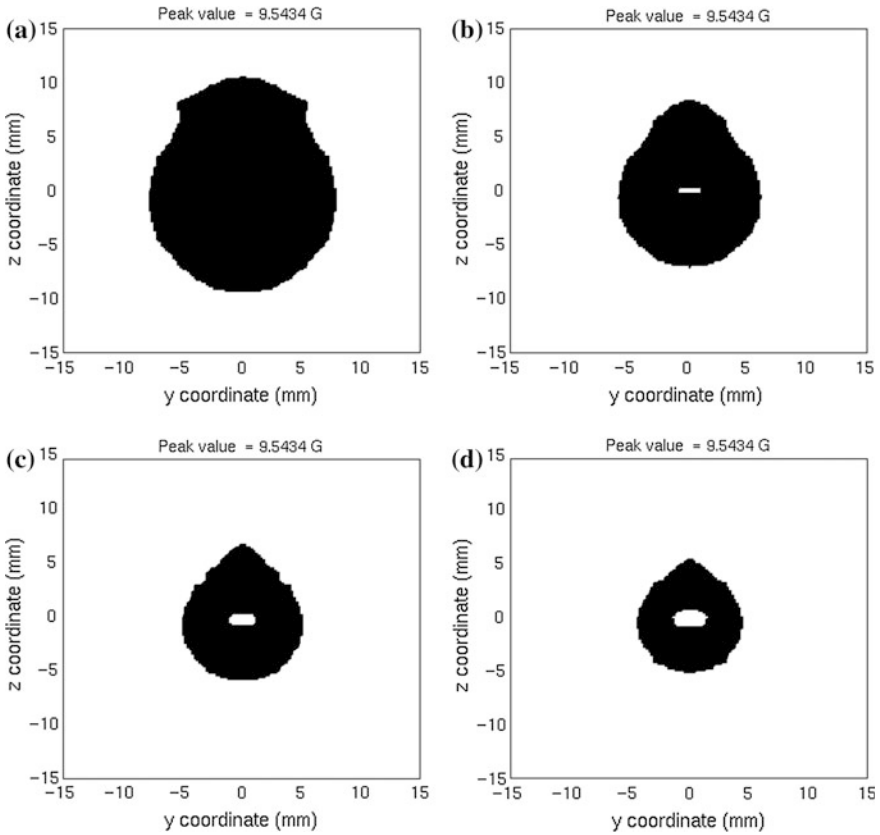
Fig. 15.41 MO images with different thresholds under infinite foil and rotating excitation current at a frequency of 1.5 kHz; crack length is 5 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 9.5434 G

15.3.8.5 Effect of Frequency

Figures 15.41, 15.42, 15.43, and 15.44 show the binary MO images obtained at various frequencies (1.5, 3, 5, and 10 kHz, respectively). With the increase of excitation frequency, the peak value of flux density at the sensor layer increases. The crack in the second layer is detectable at low frequencies (1.5 and 3 kHz), as shown in Figs. 15.41 and 15.42, and less detectable at high frequencies (5 and 10 kHz), as shown in Figs. 15.43 and 15.44, due to the skin effect. Also, lower threshold value makes the crack easier to detect. The most important effect of frequency is that as frequency increases, the threshold for defect detection at which the crack can be detected will decrease. This is verified in Figs. 15.41, 15.42, 15.43, and 15.44. In Figs. 15.41, 15.42, and 15.43 (1.5, 3, and 5 kHz), the thresholds for defect detection are 4, 3, and 2 G, respectively. In Fig. 15.44 (10 kHz), the crack is not detectable at all threshold values.
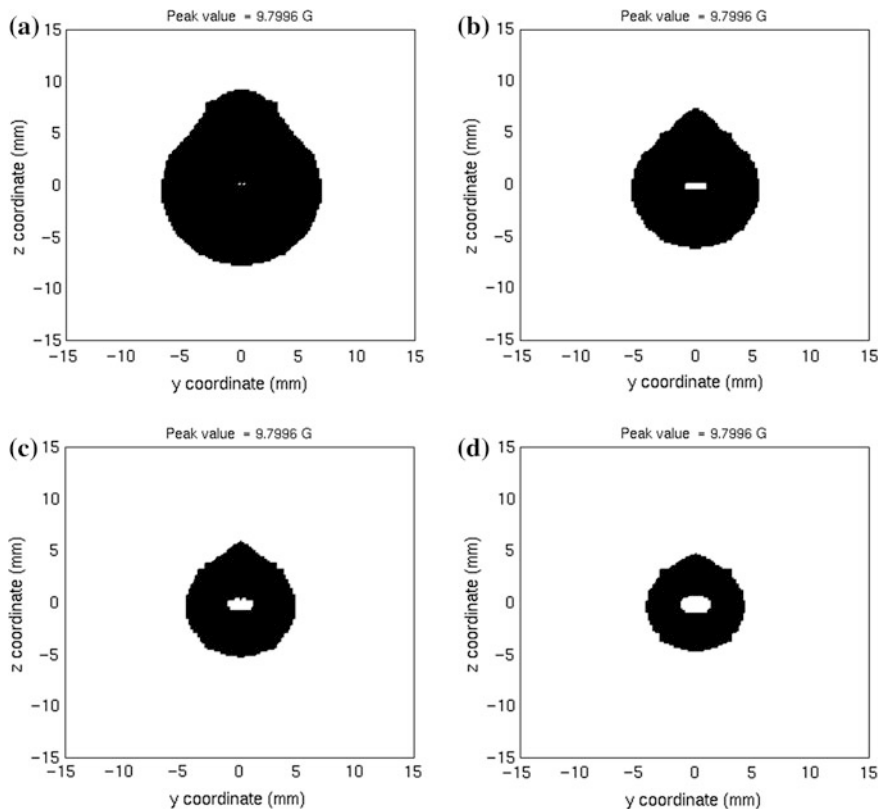
Fig. 15.42 MO images with different thresholds under infinite foil and rotating excitation current at a frequency of 3 kHz; crack length is 5 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 9.7996 G

15.3.8.6 Effect of Crack Size

Figures 15.45, 15.46, 15.47, and 15.48 show the binary MO images of rectangular cracks of lengths 3, 4, 5, and 8 mm, respectively, obtained using various thresholds. Obviously, as the length of crack increases, it is easier to detect this anomaly.

15.3.8.7 Effect of Crack Type

Figures 15.49, 15.50, 15.51, and 15.52 show the binary MO images of wedge cracks of lengths 3, 4, 5, and 8 mm, respectively. Unlike rectangular crack, the wedge crack is easy to detect only when the crack length is 8 mm (among the crack sizes studied) and the threshold is not higher than 2 G (Fig. 15.52a, b). The
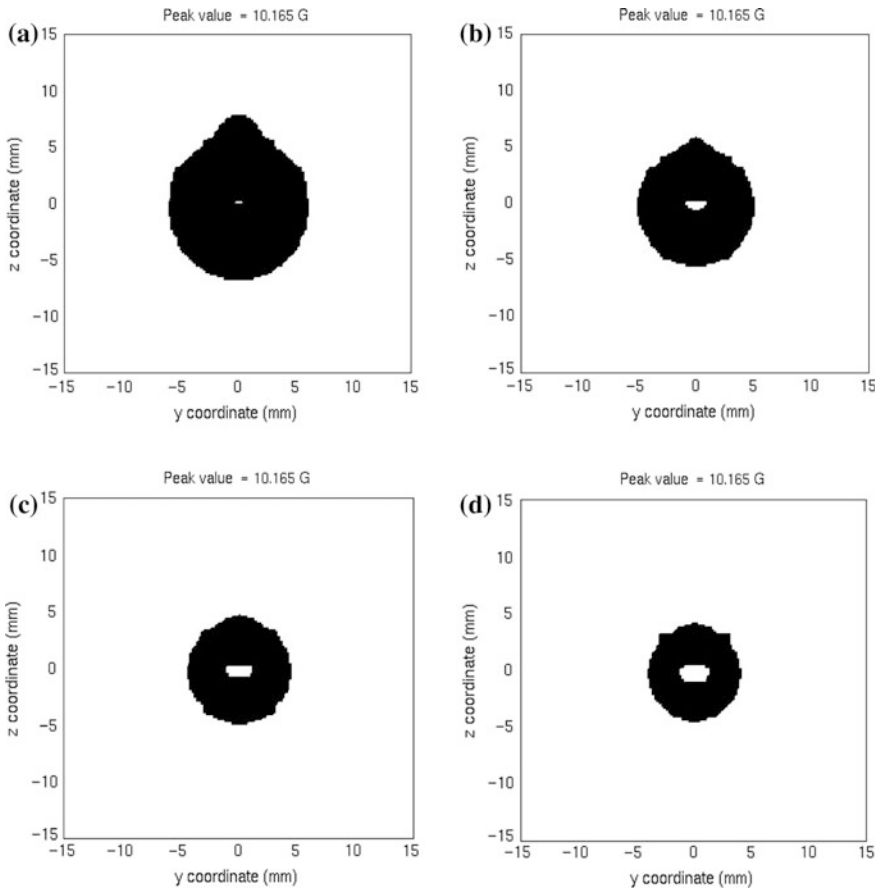
Fig. 15.43 MO images with different thresholds under infinite foil and rotating excitation current at a frequency of 5 kHz; crack length is 5 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 10.165 G

detectability of smaller cracks can be improved, and their POD can be quantified using measures such as skewness function, which will be discussed in Sect. 15.4: Data Analysis for MOI Data.

15.3.8.8 Effect of Air Gap Between Plates

Two cases have been discussed in this subsection. The first one embeds a rivet in a two-layer aluminum structure. The second case is that of a rivet and a 5 mm by 1 mm radial crack located in the second layer growing from the rivet. Rotating excitation currents with frequency of 1 kHz were applied in both cases. The air gap width was increased by moving down the second layer. The variation of the signal
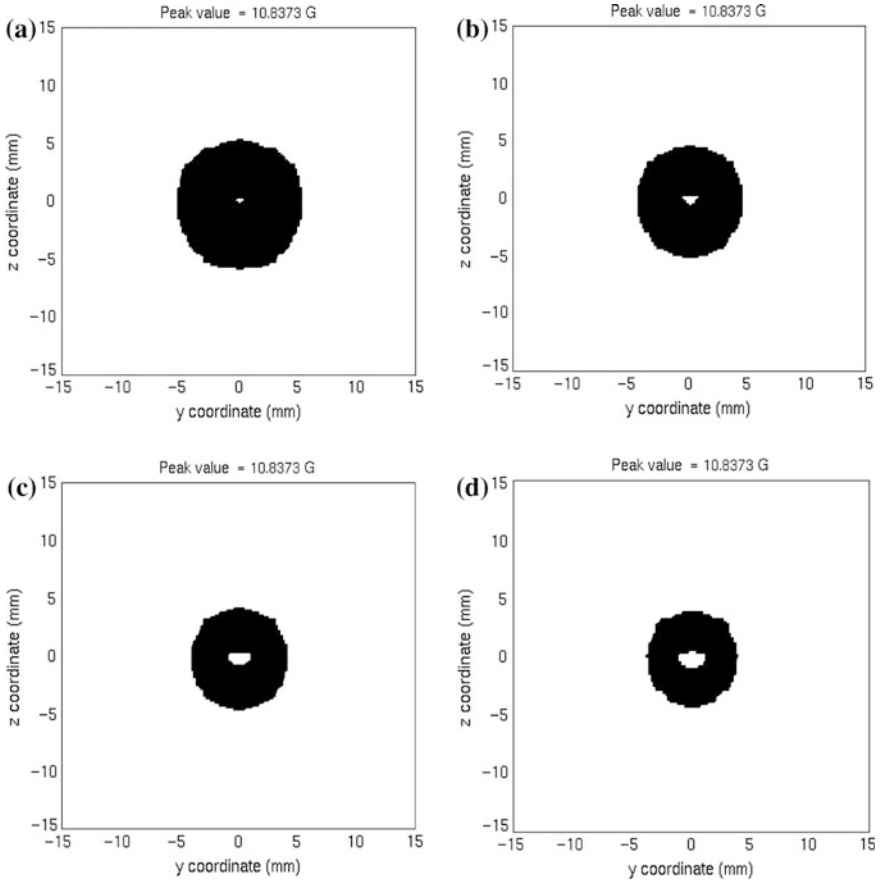
Fig. 15.44 MO images with different thresholds under infinite foil and rotating excitation current at a frequency of 10 kHz; crack length is 5 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 10.837 G

peak value versus air gap (between plates) width with and without crack is shown in Fig. 15.53. The peak value jumps up when the air gap width is increased from 0 to 0.01 mm because the presence of an air gap (even the smallest air gap) materially affects the current distribution in the plates. However, in both cases, the peak values decrease monotonically with increase in the width of air gap between the plates. The presence of an air gap is also seen to affect the pattern of MO image significantly. In Fig. 15.54a–c, a crack is shown in Fig. 15.54b, c, but not detectable in Fig. 15.54a.

We continued this study by using rotating excitation current at frequencies of 1, 5, and 50 kHz. The relationships between the signal peak value and air gap width at different frequencies are shown in Fig. 15.55. All the curves show a jump in peak value when the air gap width was increased from 0 to 0.01 mm. Afterward, the peak
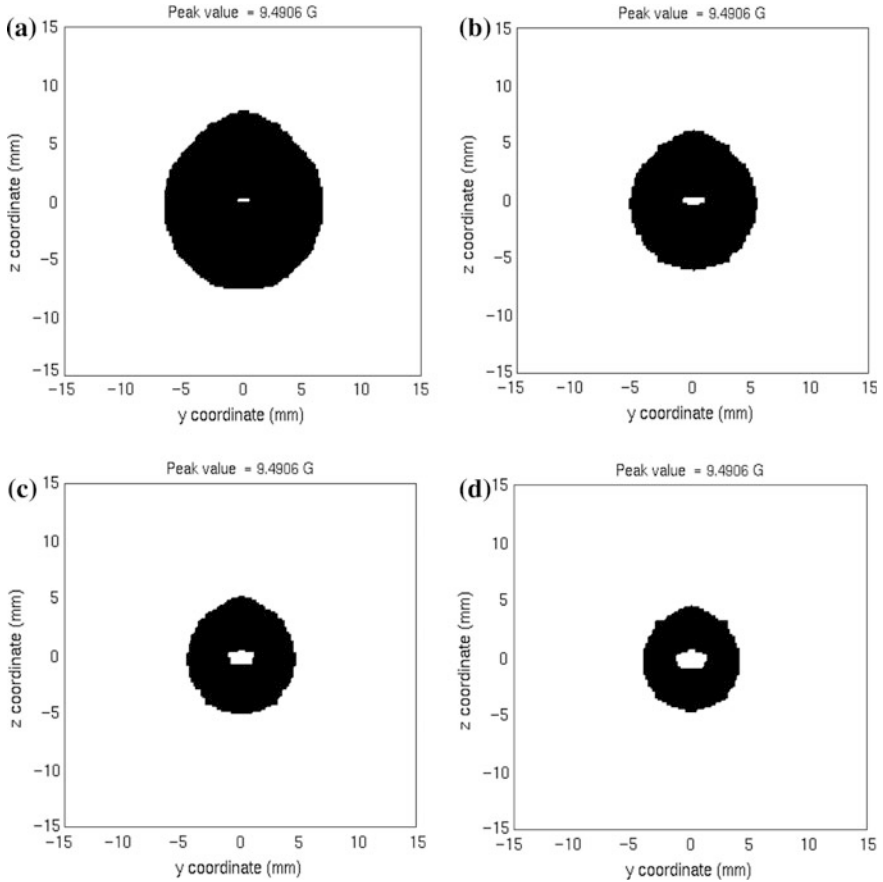
Fig. 15.45 MO images with different thresholds under infinite foil and rotating excitation current at a frequency of 3 kHz; crack length is 3 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 9.4906 G

value of the signal at 1 kHz decays linearly with the increase in the air gap width. In contrast, the signal value at 5 kHz increases with increasing air gap width. Little variation is observed at 50 kHz.

In order to investigate the reason for the jump in peak value when the air gap width is increased from 0 to a small value (0.01 mm), we replaced the air gap by some material whose conductivity was gradually changed to study the relation of peak value versus conductivity. The conductivity was changed from the value of aluminum (which is equivalent to no air gap) to 0 (which is equivalent to with air gap). The frequency is fixed at 5 kHz. The relationship of signal peak value versus conductivity at the air gap region is shown in Fig. 15.56. Obviously, when the conductivity is small, reducing the conductivity increases the peak value dramatically. This fact explains the reason for the jump in peak value when the air gap
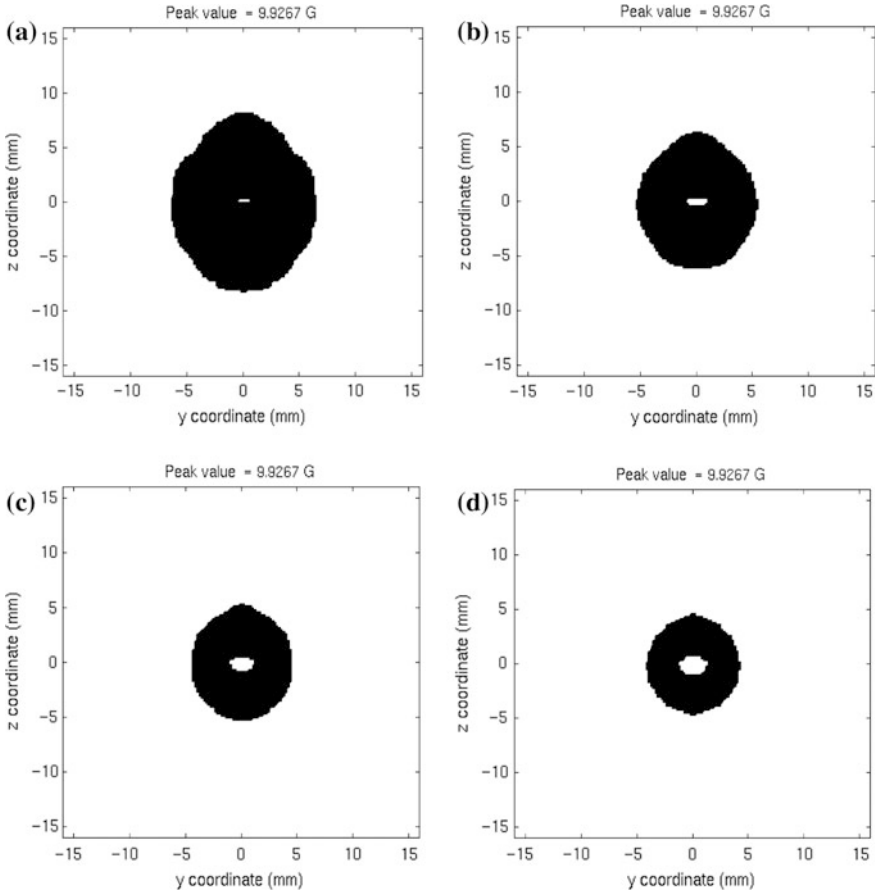
Fig. 15.46 MO images with different thresholds under infinite foil and rotating excitation current at a frequency of 3 kHz; crack length is 4 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 9.9267 G

width is increased from 0 to 0.01 mm. Another good review on MOI parametric study can refer to [29].

15.4 Data Analysis for MOI Data

15.4.1 MO Image Post-processing, Enhancement, and Automatic Detection

The overall approach for automated MO rivet hole image classification is depicted in Fig. 15.57. The raw image obtained from the MOI acquisition system is applied
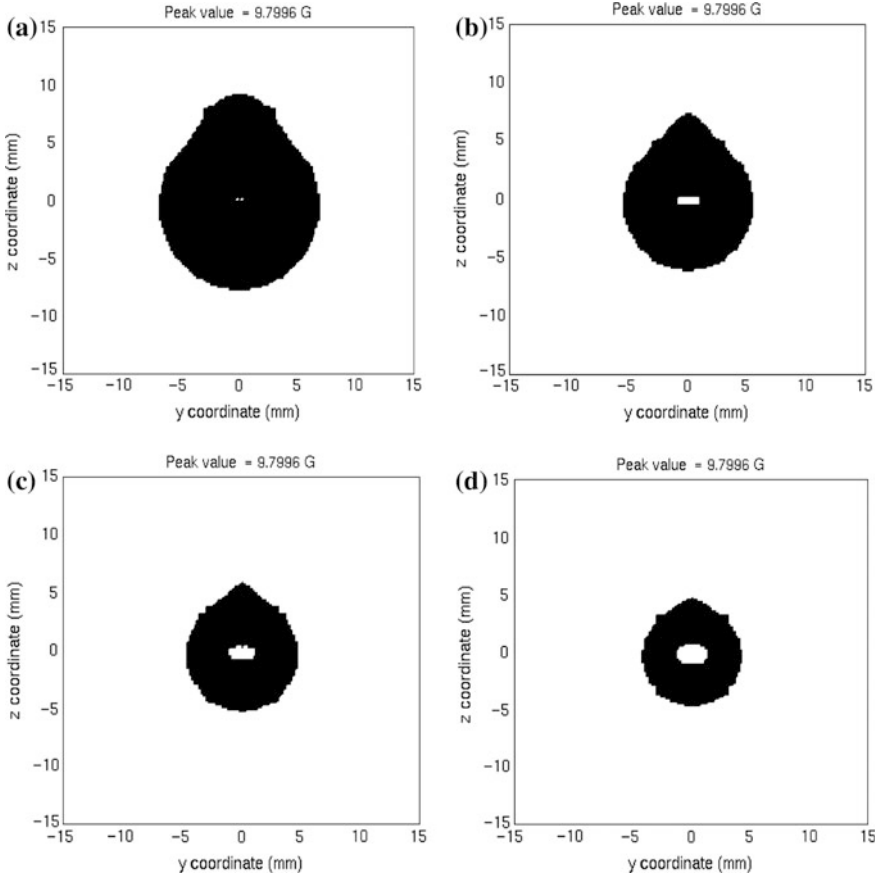
Fig. 15.47 MO images with different thresholds under infinite foil and rotating excitation current at a frequency of 3 kHz; crack length is 5 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 9.7996 G

to the motion-based filter [34] to remove the background noise associated with domain structures in the MO sensor [13]. The filtered image is enhanced and thresholded to obtain a binary image, which is used in the subsequent rivet hole detection and classification modules.

The detection and classification of a rivet hole is based on the skewness value associated with the binary image. The rivet center detection is performed using one of two different approaches, namely Hough transformation and morphological image processing [35]. All the classification results in this section are obtained using the morphological image processing operations of dilation and erosion with varying structural elements [36]. Rivet quantification is first performed using skewness feature of the data. The ideal rivet hole image is circular, and hence, its skewness must be zero. If the computed skewness value is greater than some tolerance, the rivet hole is classified as defective; otherwise, the rivet site is deemed good.
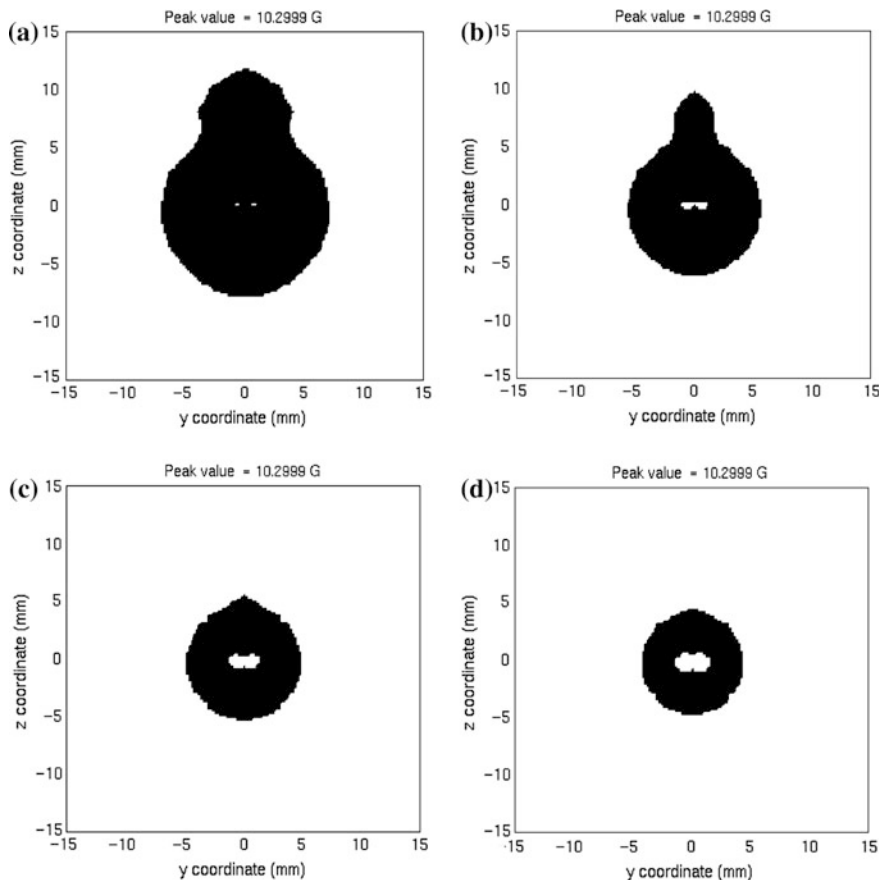
Fig. 15.48 MO images with different thresholds under infinite foil and rotating excitation current at a frequency of 3 kHz; crack length is 8 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 13.000 G

15.4.2 Characterization of MO Images

The skewness of MO image of rivet hole and its quantification play a critical role in the automatic rivet hole classification [37]. The skewness function should capture the deviation from circularity of the rivet hole image, be robust under small variations in the measured image, and at the same time have discriminatory information between good and defective rivet holes. Several definitions of the skewness associated with an MO image have been investigated [38, 39] and discussed in the following sections.

Skewness function S_1: The first definition studied was designed to be independent of the excitation frequency and is expressed as follows:
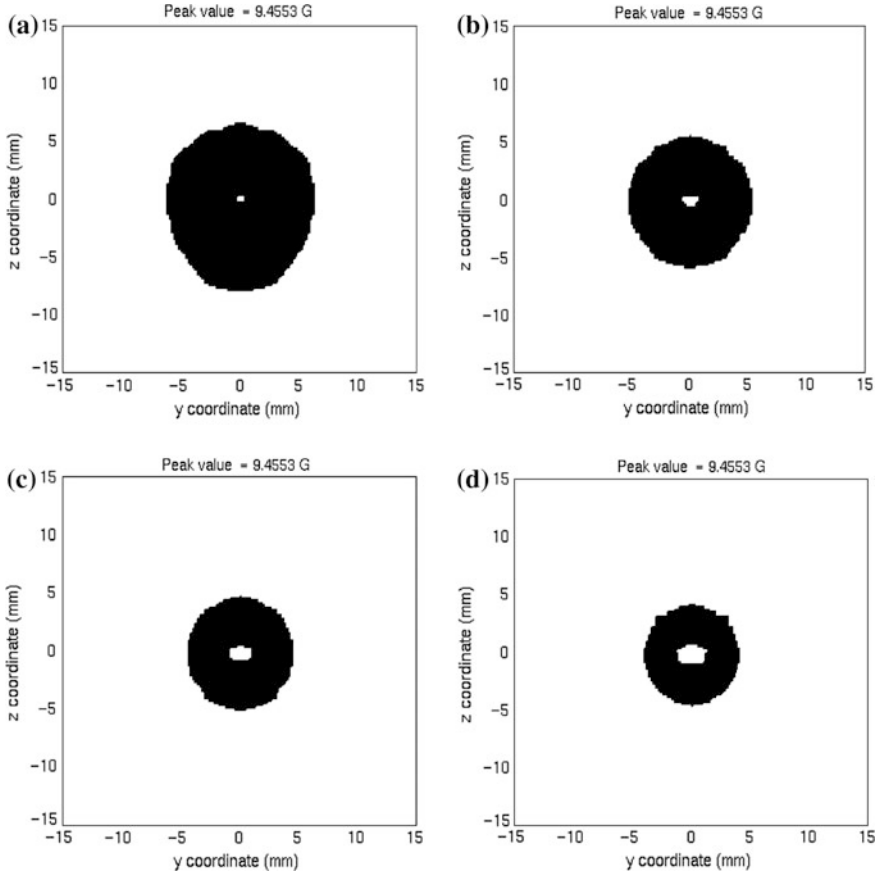
Fig. 15.49 MO images with wedge crack and different thresholds under infinite foil and rotating excitation current at a frequency of 3 kHz; crack length is 3 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 9.4553 G

$$S_1 = \frac{f^{1/3}}{100} \frac{R}{R+B} \frac{\sum_i (D_i - r)^2}{r^2 + \sum_i (D_i - r)^2} \tag{15.21}$$

where f is the frequency parameter which is 50 kHz for the surface defects and 1.5, 3, or 5 kHz for subsurface defects. Here, B is the "width" of the region outside the rivet in the image, r is the average measured radius of rivet, and $\{D_i, i = 1, 2, \ldots, n\}$ are the distances from the rivet center to the edge pixels. Finally, R is computed by calculating the histogram of D_i and picked at its peak location. All these parameters are illustrated in Fig. 15.58.

Skewness function S_2: The second definition of skewness function that was studied is based on the autocovariance function, $\text{cov}(D_i, D_j)$ of the series $\{D_i\}$. The
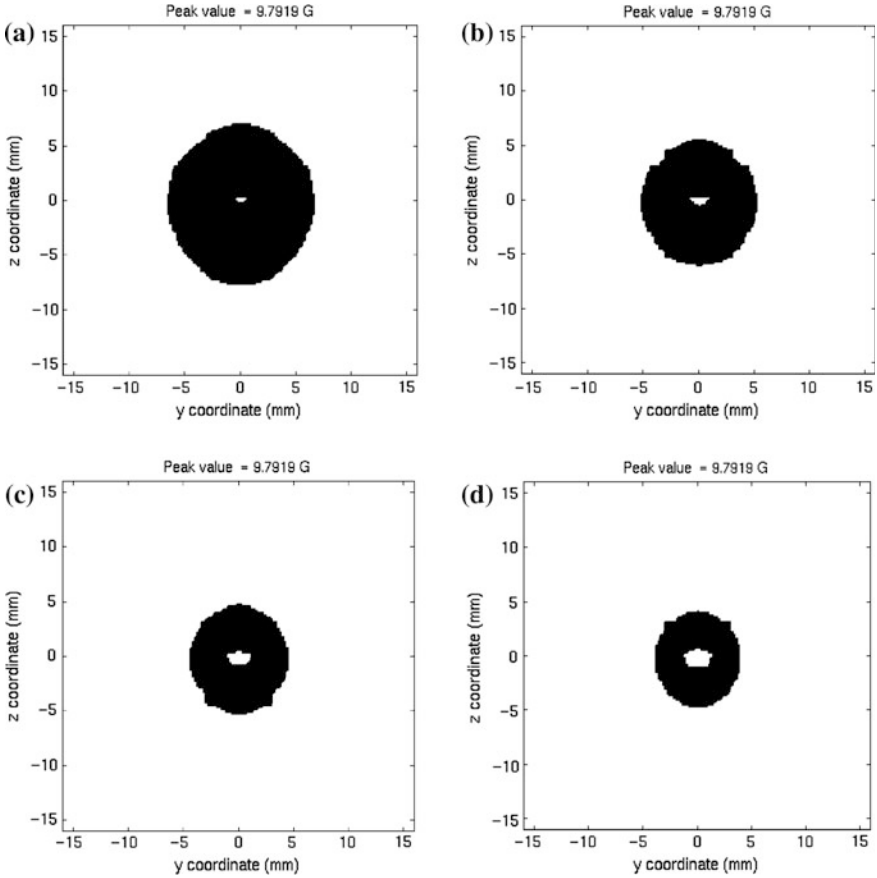
Fig. 15.50 MO images with wedge crack and different thresholds under infinite foil and rotating excitation current at a frequency of 3 kHz; crack length is 4 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 9.7919 G

vector-valued skewness function is derived from features in the autocovariance plot and is hence defined in a two- or higher-dimensional space for improved separability between classes. Consider a set of n distances denoted by $\{D_1, D_2, \ldots, D_n\}$, as illustrated in Fig. 15.58b in which the circle represents the boundary of the binary MO image of a rivet hole. The continuous form of normalized covariance definition is as follows:

$$\mathrm{cov}(\theta') = \frac{\int_0^{2\pi} (D(\theta) - r)(D(\theta + \theta') - r)\mathrm{d}\theta}{\int_0^{2\pi} D(\theta)^2 \mathrm{d}\theta} \tag{15.22}$$

The corresponding discrete form of normalized covariance definition is given by:
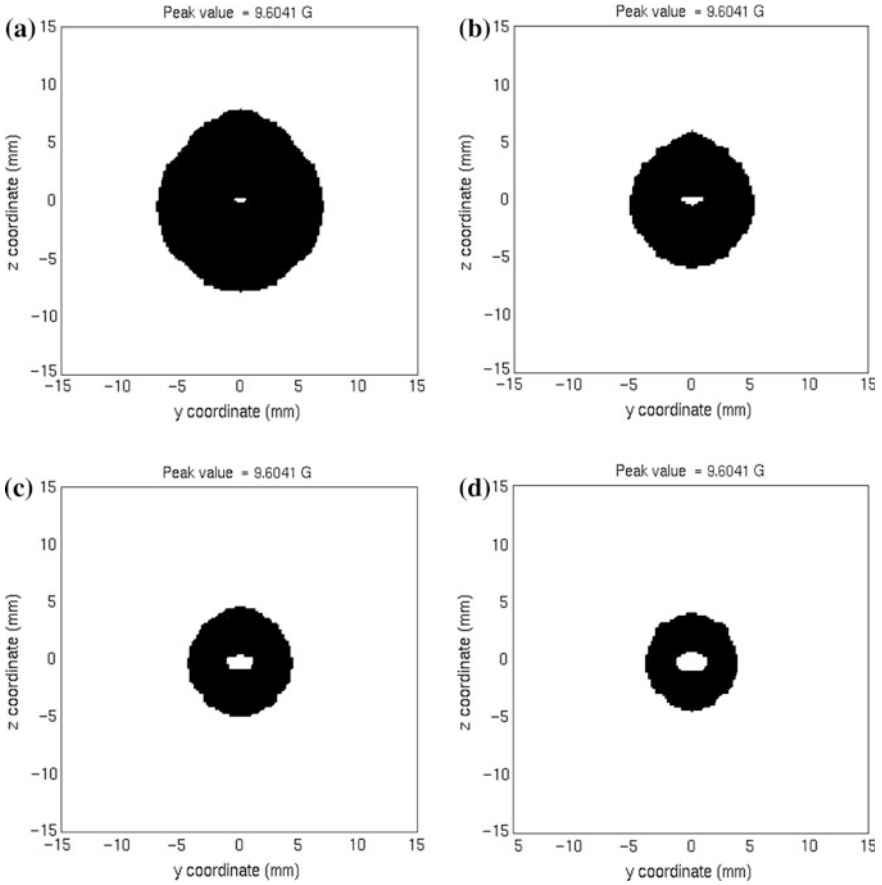
Fig. 15.51 MO images with wedge crack and different thresholds under infinite foil and rotating excitation current at a frequency of 3 kHz; crack length is 5 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 9.6041 G

$$\text{Cov}(\theta_k) = \frac{\sum_{i=1}^{n} (D(\theta_i) - r) \cdot (D(\theta_i + \theta_k) - r)\delta\theta_i}{\sum_{i=1}^{n} D^2(\theta_i)\delta\theta_i} \tag{15.23}$$

where $\theta_k = \theta_{k-1} + \delta\theta_k$, $\delta\theta_i$ is the incremental angle between D_{i+1} and D_i, and the initial angle θ_0 is set to 0°. Typical autocovariance curves of the series $\{D_i\}$, derived from MO images, are shown in Fig. 15.59. Three skewness parameters are derived from the autocovariance function, namely central peak ratio (CPR), noise-to-signal ratio (NSR), and central peak width (CPW), as shown in Fig. 15.60. CPR is the ratio of the central peak height H_{CP} to the total height of the covariance curve H_{T}, defined as follows:
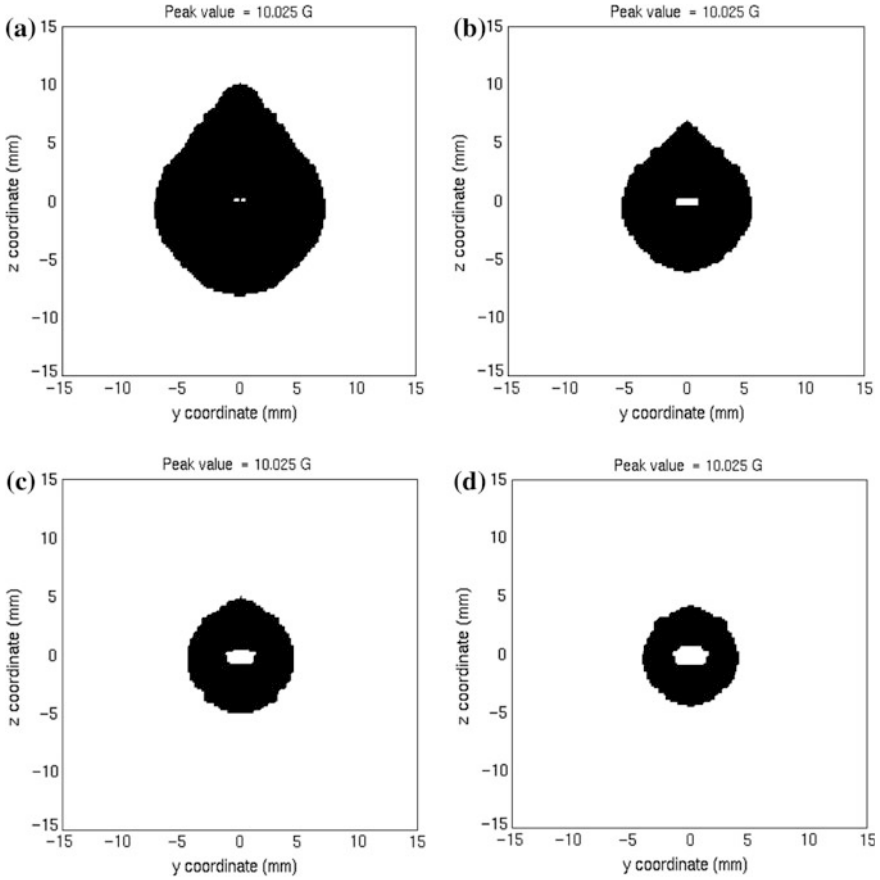
Fig. 15.52 MO images with wedge crack and different thresholds under infinite foil and rotating excitation current at a frequency of 3 kHz; crack length is 8 mm; thresholds are as follows: **a** 1 G, **b** 2 G, **c** 3 G, and **d** 4 G, peak value of normal magnetic flux density being 10.025 G

$$\text{CPR} = \frac{H_{CP}}{H_T} \tag{15.24}$$

CPW is the width of the central peak at its 50 % peak value. NSR is the noise-to-signal ratio defined from continuous form of normalized covariance definition:

$$\text{NSR} = \frac{\int_0^{2\pi} (D(\theta) - r)(D(\theta) - r)\mathrm{d}\theta}{\int_0^{2\pi} D^2(\theta)\mathrm{d}\theta} \tag{15.25}$$

which is also the starting point of the covariance curve cov(0). All parameters are illustrated in Fig. 15.60. Using these skewness parameters, a 2D feature space spanned by CPR and NSR or a 3D feature space spanned by CPR, CPW, and NSR
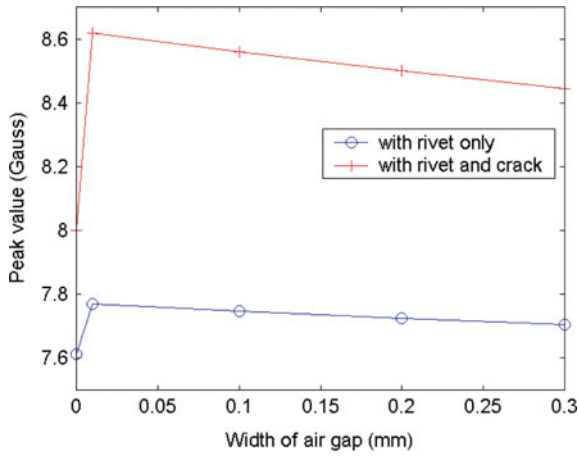
Fig. 15.53 Relations of signal peak value versus air gap width with and without crack
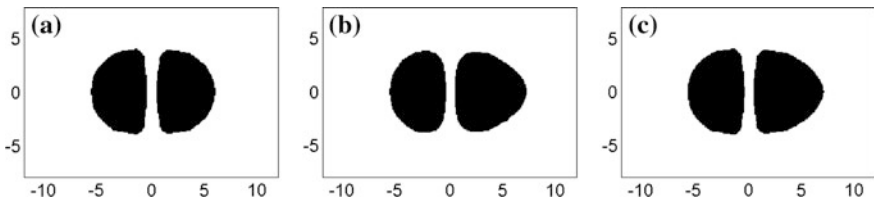
Fig. 15.54 Binary MO images with rivet and a 5-mm crack at the bottom layer, with air gap width: **a** 0, **b** 0.01 mm, and **c** 0.1 mm
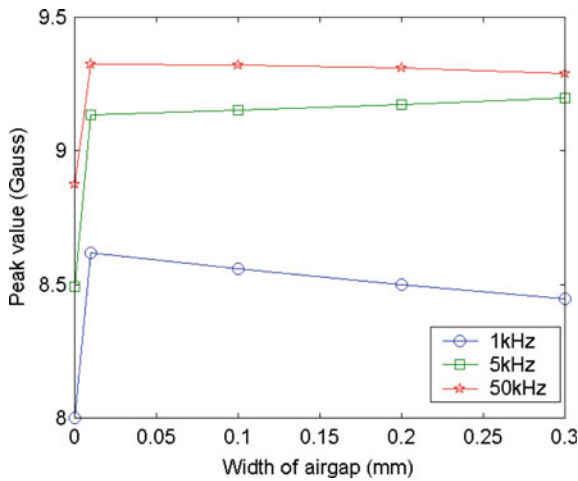
Fig. 15.55 Relations of signal peak value versus air gap width at different frequencies

Fig. 15.56 Peak value of normal magnetic flux density versus conductivity at "air gap region": **a** normal plot and **b** logarithm plot
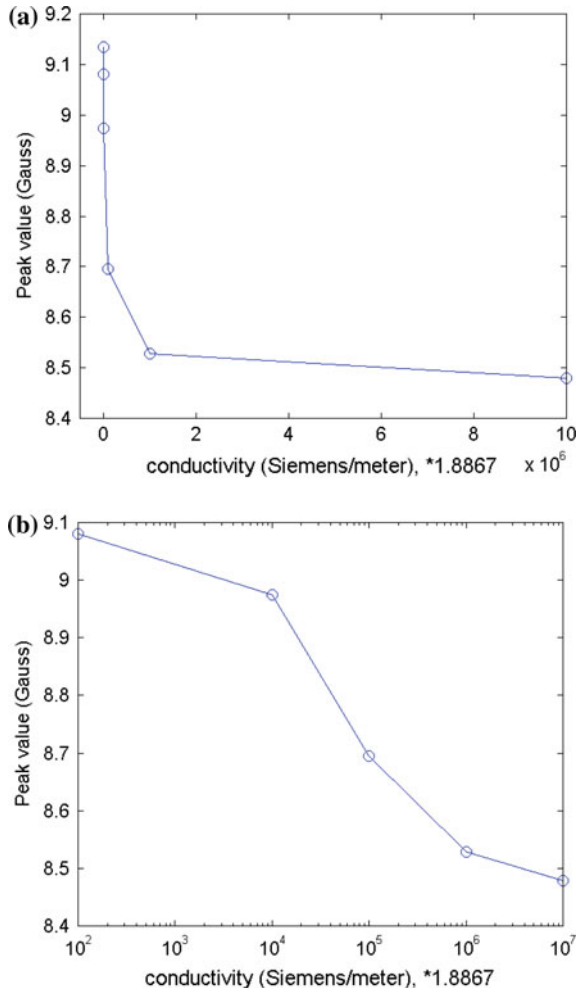
can be constructed, where the coordinates in 2D space are (CPR, NSR) and in 3D space are (CPR, CPW, NSR). In the 2D feature space, according to the definitions above and characteristic of the binary MO images, it is seen that for small defects, the value of CPR is close to 1.0; also, the value of NSR is proportional to the size (radial length) of the crack. The skewness vector values in this case are defined based on the multidimensional coordinates,

$$S_2 = \begin{cases} (\text{CPR}, \text{NSR}) & \text{for 2D space} \\ (\text{CPR}, \text{NSR}, \text{CPW}) & \text{for 3D space} \end{cases} \tag{15.26}$$

Skewness function S_3: A third definition of skewness is based on the normalized central moments of 2D binary-valued rivet hole images. For a 2D digital image function $I(x, y)$, the discrete central moment can be expressed as follows:
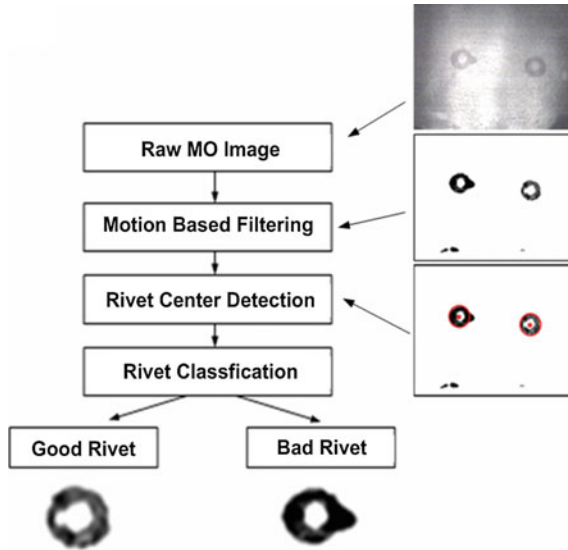
Fig. 15.57 Overall approach of automated rivet classification algorithm
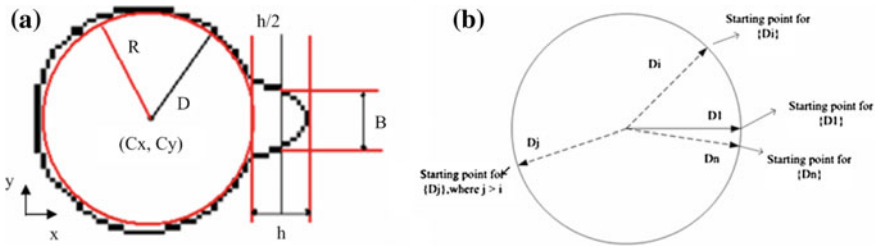
Fig. 15.58 Illustration of parameters in skewness function S_1 **a** the parameters and **b** the set of distances D_i from the binary rivet hole images

$$\mu_{pq} = \sum_{x} \sum_{y} (x_i - \bar{x}_c)^p (y_j - \bar{y}_c)^q I(x_i, y_i) \qquad (15.27)$$

where $(\bar{x}_c, \bar{y}_c)$ is the center of the MO rivet image obtained using the morphological image processing operators of dilation and erosion. For the binary-valued rivet images after enhancement and thresholding, $I(x, y)$ is simplified to

$$I(x, y) = \begin{cases} 1 & (x_i, y_j) \text{ is image pixels} \\ 0 & \text{otherwise} \end{cases} \qquad (15.28)$$

The normalized central moments, denoted by M_{pq}, are then computed as follows:
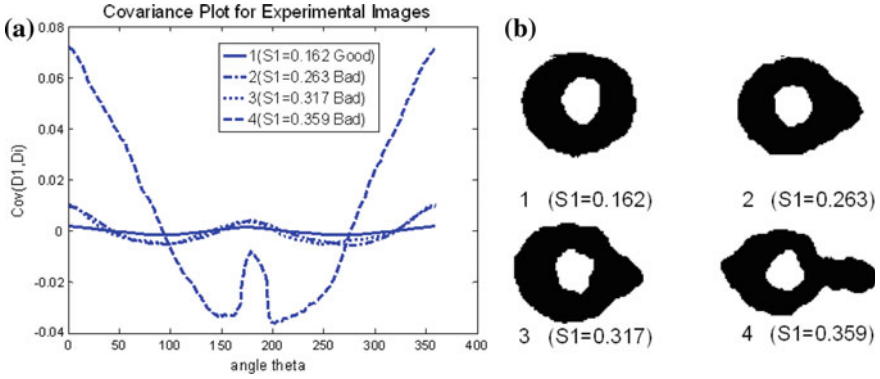
Fig. 15.59 Autocovariance functions for good and bad rivet holes. **a** Covariance curves with different S values. **b** Corresponding binary images
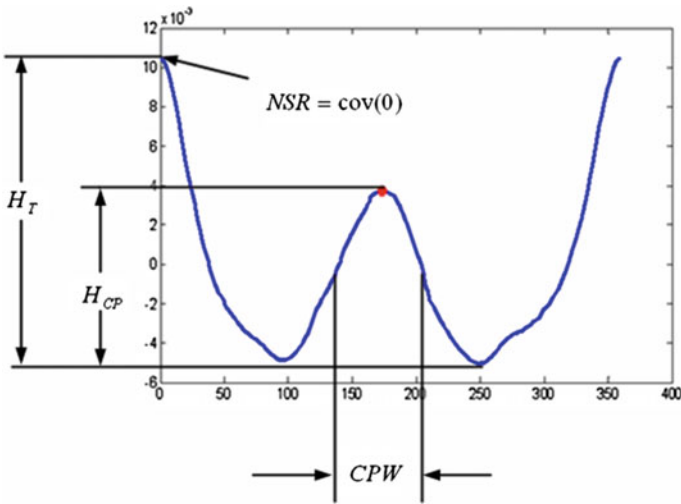
Fig. 15.60 Skewness features derived from the covariance curve

$$M_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \qquad (15.29)$$

where $\gamma = \frac{p+q}{2} + 1$, for $p + q = 2, 3, \ldots$.

The skewness function S_3 was chosen from the various normalized central moments, based on their discriminatory abilities, as the vector S_3 in (15.30)

$$S_3 = \{M_{20}, M_{22}, M_{40}\} \qquad (15.30)$$
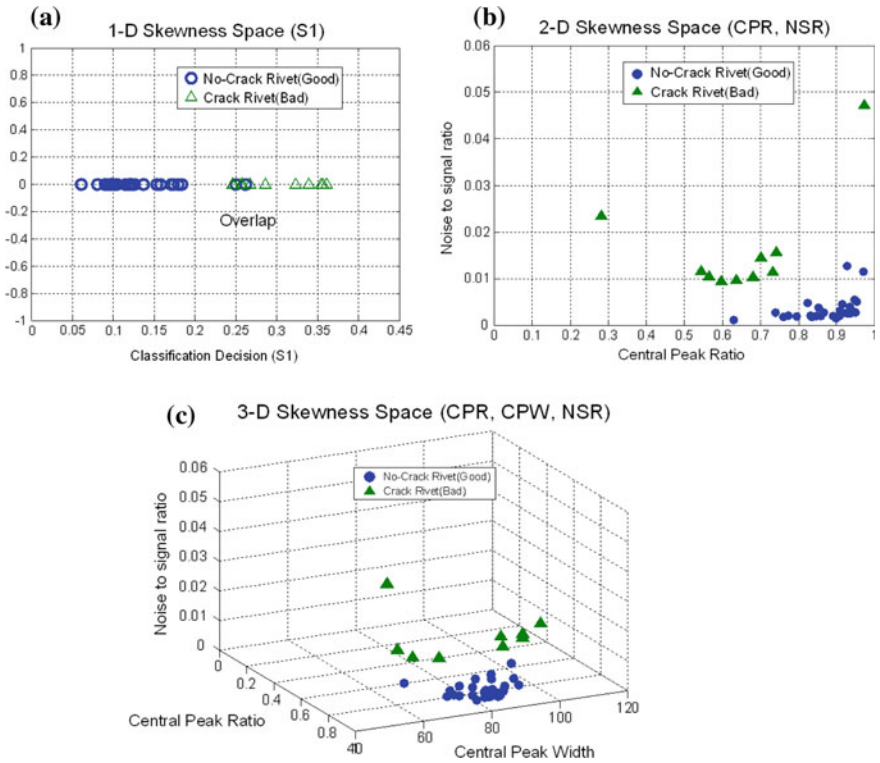
Fig. 15.61 Classification results in multidimensional skewness space: **a** 1D space, **b** 2D space, and **c** 3D space
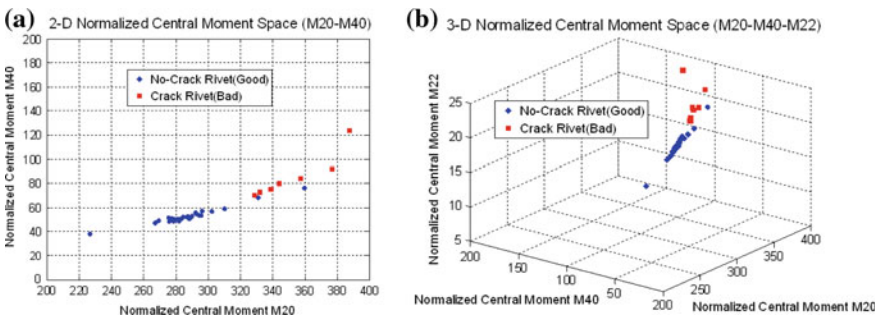
Fig. 15.62 Classification results in normalized central moment space: **a** (M_{20}, M_{40}) space and **b** (M_{20}, M_{40}, M_{22}) space

Results of classification using these skewness criteria are presented in Figs. 15.61 and 15.62.

15.4.3   Probability of Detection Studies

Probability of detection (POD) is used in NDE as a method for quantitative assessment of system performance. The value of POD of a particular flaw of a given dimension, orientation, and location using a given measurement sensor/system can be determined by generating conditional PDF of the measurement signal. In NDE inspection, accept/reject decisions are made by choosing a threshold (Th) such that all signals above the threshold will be classified as flaw signals and signals below the threshold will be interpreted as noise. For the binary MO images, the definitions of POD and probability of false alarm (PFA) can be simplified as follows:

$$\text{POD} = \frac{\sum_{i=1}^{N} \alpha_i}{N} \times 100\,\% \tag{15.31}$$

$$\text{PFA} = \frac{\sum_{i=1}^{N} \alpha_i^*}{N^*} \times 100\,\% \tag{15.32}$$

where N and N^* are the total number of images corresponding to flawed and good rivet holes, respectively, and α_i and α_i^* are the numbers of corresponding classification decisions of the automated rivet hole classification system. $\alpha_i = 1$ implies that the rivet hole is classified as flawed, whereas $\alpha_i = 0$ means it is a good rivet hole. When the value of S_1 is greater than or equal to the classification threshold, the rivet site was classified as having a crack. POD plots with varying classification threshold values are shown in Fig. 15.63. A systematic POD study of MOI data can be found in the latest literature [40] (Fig. 15.64).

15.5   Summary

This chapter gives a comprehensive review on the new sensing technology, magneto-optic imaging for NDE, and discusses the recent advances on MOI including the development of novel optical sources and magnetic film materials for various applications, the integration of MOI sensors with automated systems, and the state-of-the-art post-processing algorithms.
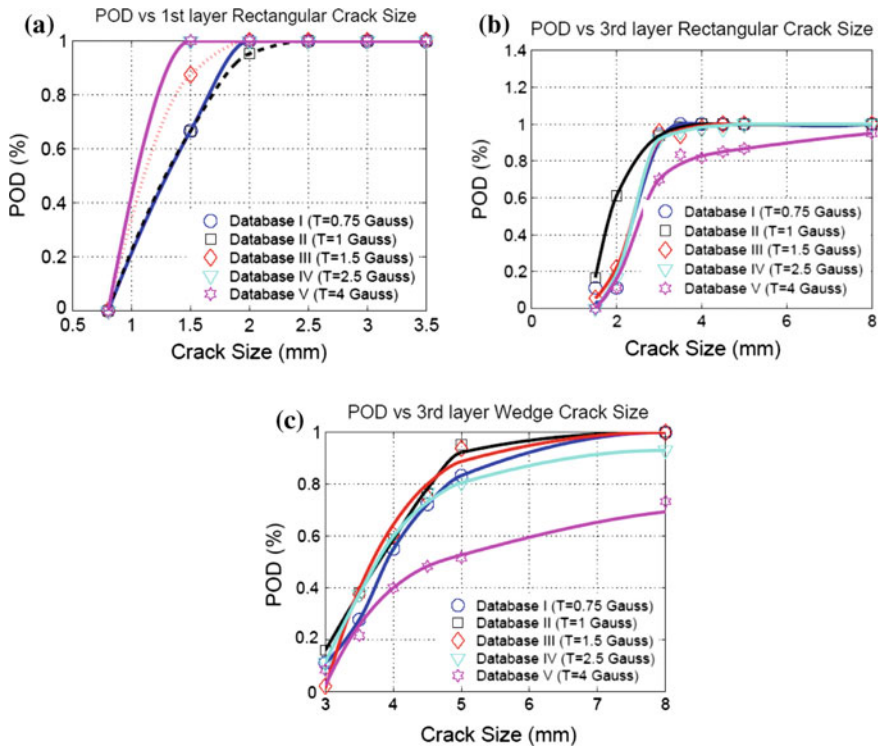
Fig. 15.63 POD versus crack size. **a** 1st layer rectangular crack with PFA = 0.0516, **b** 3rd layer rectangular crack with PFA = 0.1528, and **c** POD versus 3rd layer wedge crack with PFA = 0.0321

Fig. 15.64 POD versus skewness for 1st layer rectangular crack (*solid line*) and 3rd layer rectangular crack (*dashed line*) corresponding to the database with threshold = 1.0 Gauss
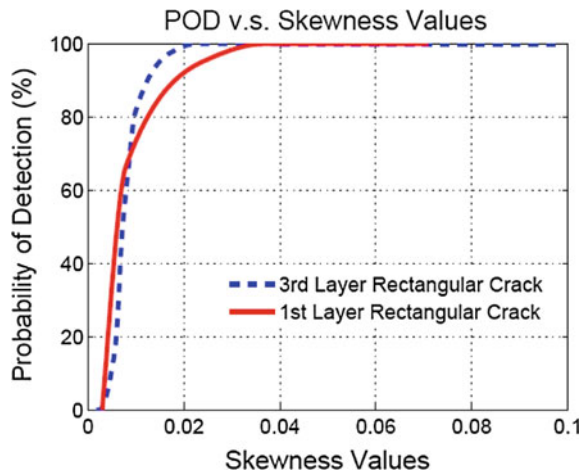
References

1. Fitzpatrick GL, Thome DK, Skaugset RL, Shih EYC, Shih WCL (1993) Magneto-optic/eddy current imaging of aging aircraft: a new NDI technique. Mater Eval 51(12):1402–1407

2. Fitzpatrick GL et al (1993) Novel eddy current field modulation of magneto-optic garnet films for real-time imaging of fatigue cracks and hidden corrosion. In: Proceedings of SPIE 2001, nondestructive inspection of aging aircraft. doi:10.1117/12.163845

3. Udpa SS, Moore PO (2004) Nondestructive testing handbook, third edition: electromagnetic testing, vol 5. Amer Society for Nondestructive

4. Morozov M, Novotny P (2002) Evaluation of eddy current probes based on local field excitation. NDT&E Int 35:147–153

5. Gomez RD, Kratz AR, Burke ER, Mayergoyz ID (1995) Spatially-resolved magnetic imaging of magneto-optic recording. IEEE Trans Magn 31(6):3265–3267

6. Dimonte G (1992) A magnetooptic imaging probe for continuous magnetic field profiles. Rev Sci Instrum 63:5151–5153

7. Muller H, Suenaga M, Yokoyama Y (1991) Magneto-optic images of flux distribution and creep in $YBa_2Cu_3O_7$. J Appl Phys 70:4409–4414

8. Thome DK, Fitzpatrick GL, Skaugset RL, Shih WCL (1996) Aircraft corrosion and crack inspection using advanced magneto-optic imaging technology. In: SPIE

9. Didosyan YS, Hauser H, Nicolics J (2000) Magneto-optical current sensors of high bandwidth. Sens Actuators 81:263–267

10. Cheng Y, Deng Y, Bai L, Chen K (2013) Enhanced laser-based magneto-optic imaging system for nondestructive evaluation applications. IEEE Trans Instrum Meas 62(5):1192–1198

11. Cheng YH, Zhou ZF, Tian GY (2007) Enhanced magneto-optic imaging system for nondestructive evaluation. NDT&E Int 40:374–377

12. Agalidi Y, Kozhukhar P, Levyi S, Rogozhinsky S, Shumsky I (2012) Eddy current fields/magnetic recording/magneto-optic imaging NDI method. Nondestruct Test Eval 27(2):109–119

13. Ferrari H, Bekeries V, Thibeault M, Johansen TH (2007) Magneto-optic imaging: normal and parallel field components of in-plane magnetized samples. J Magn Magn Mater 313:98–106

14. Heidmann J (2009) Magneto-optic write-head characterization using a garnet imaging film. IEEE Trans Magn 45(10):3648–3651

15. Johansen TH et al (1996) Direct observation of the current distribution in thin superconducting strips using magneto-optic imaging. Phys Rev B 54(22):264–269

16. Novotny P et al (2002) On the application of magneto-optic films in NDE. NDT&E Int 35:227–232

17. Novotny P (2004) A magneto-optic imager for NDT applications. NDT&E Int 37:645–649

18. Tehranchi MM et al (2011) The inspection of magnetic flux leakage from metal surface cracks by magneto-optical sensors. Sens Actuators A: Phys 172:365–368

19. Vishnevskii V, Berzhansky V, Mikhailov V, Pankov F, Nedviga A, Nesteruk A et al (2012) In: International conference on oxide materials for electronic engineering

20. Cheng Y, Deng Y, Cao J, Xiong X, Bai L, Li Z (2013) Multi-wave and hybrid imaging techniques: a new direction for nondestructive testing and structural health monitoring. Sensors 13(12):16146–16190

21. Joubert PY, Pinassaud J (2006) Linear magneto-optic imager for non-destructive evaluation. Sens Actuators A 129:126–130

22. Baziljevich M, Barness D, Sinvani M et al (2012) Magneto-optical system for high speed real time imaging. Rev Sci Instrum 83(8):083707

23. Murakami H, Tonouchi M (2010) High-sensitive scanning laser magneto-optical imaging system. Rev Sci Instrum 81:013701

24. Gao X, Liu Y, You D (2014) Detection of micro-weld joint by magneto-optical imaging. Opt Laser Technol 62:141–151

25. Pan Z et al (2013) A novel defect inspection method for semiconductor wafer based on magneto-optic imaging. J Low Temp Phys 170:436–441
26. Mohammad Nur-E-Alam M, Vasiliev M, Kotov VA, Alameh K (2011) Highly bismuth-substituted, record-performance magneto-optic garnet materials with low coercivity for applications in integrated optics, photonic crystals, imaging and sensing. Opt Mater Express 1:413–427
27. Zhu Y (2005) Modern techniques for characterizating magnetic materials. Kluwer Academic Publishers
28. Xuan L (2002) Finite element and meshless methods in NDT applications. In: ECE, Iowa State University
29. Zeng Z, Liu X, Deng Y, Udpa L, Xuan L, Shih WCL, Fitzpatrick GL et al (2006) A parametric study of magneto-optic imaging using finite-element analysis applied to aircraft rivet site inspection. IEEE Trans Magn 42(11):3737–3744
30. Le M, Lee J, Shoji T (2011) A simulation of magneto-optical eddy current imaging. NDT&E Int 44:783–788
31. Xuan L et al (2000) Finite element model for MOI applications using A-v formulation. Rev Prog Quant Nondestruct Eval
32. Chen C (2000) Finite element modeling of MOI for NDE applications. In: ECE, Iowa State University
33. Freund RW (1992) Transpose-free quasi-minimal residual methods for non-hermitian linear systems. In: Advances in computer methods for partial differential equations. IMACS
34. Park U, Udpa L, Stockman GC (2004) Motion-based filtering of magneto-optic imagers. Image Vis Comput 22:243–249
35. Ramuhalli P, Yuan F, Park U, Slade J, Udpa L et al (2003) Enhancement of magneto-optic images. In: The 9th international workshop on electromagnetic nondestructive evaluation, pp 199–205
36. Park U, Udpa L, Stockman GC, Shih W, Fitzpatrick J (2003) Real-time implementation of motion-based filtering in magneto-optic imager. In: AIP proceedings on review of progress in quantitative nondestructive evaluation
37. Zeng Z, Udpa L, Udpa SS, Xuan L, Shih W (2008) Optimization of test parameters for magneto-optic imaging using Taguchi's parameter design and response-model approach. Res Nondestruct Eval 19(3):164–180
38. Deng Y, Liu X, Fan Y, Zeng Z, Udpa L, Shih W (2006) Characterization of magneto-optic imaging data for aircraft inspection. IEEE Trans Magn 42(10):3228–3230
39. Deng Y, Zeng Z et al (2007) Automatic classification of MOI for aircraft rivet inspection. Int J Appl Electromagn Mech 25:375–382
40. Deng Y, Liu X, Udpa L (2012) Magneto-optic imaging for aircraft skins inspection: a probability of detection study of simulated and experimental image data. IEEE Trans Reliab 61(4):901–908

Index