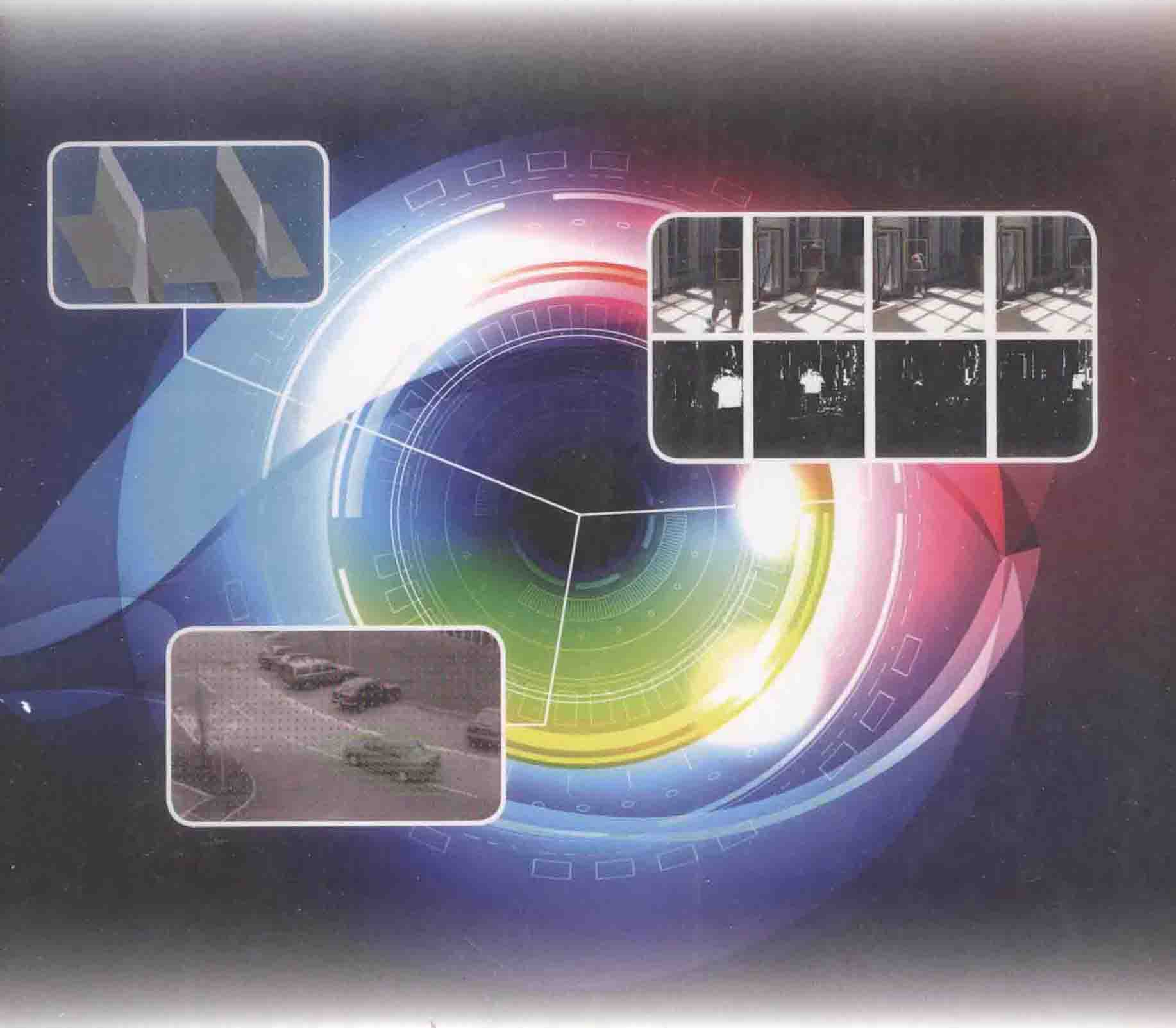
Kenneth Dawson-Howe



A Practical Introduction to COMPUTER VISION WITH OPENCV

WILEY

A PRACTICAL INTRO[illegible]N TO COMPU[illegible]SION WITH OPENCV

Kenneth Dawson-Howe
Trinity College Dublin, Ireland

WILEY

This edition first published 2014

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data applied for.

ISBN: 9781118848456

Set in 10/12pt Times by Aptara Inc., New Delhi, India
Printed and bound in Singapore by Markono Print Media Pte Ltd

1 2014

A PRACTICAL INTRODUCTION TO COMPUTER VISION WITH OPENCV

I am grateful to many people for their help and support during the writing of this book. The biggest thanks must go to my wife Jane, my children, William and Susie, and my parents, all of whose encouragement has been unstinting.

I must express my thanks to my students for their interest and enthusiasm in this subject. It is always refreshing to hear students discussing how to solve vision problems in tutorials and great to hear their solutions to problems which are often different (and sometimes better) than my own.

I thank my colleagues (in particular Arthur Hughes, Jeremy Jones and Hilary McDonald) for their encouragement and support.

Preface

Perception is essential in order for any entity to interact in a meaningful way with its environment. Humans draw on many senses (such as sight, sound, touch and smell) to perceive the world. Most machines can only receive input through simple input devices, such as keyboards and mice, or through wired and wireless communication channels. However, in recent years, cameras and microphones have been added as standard parts of computers and mobile devices (such as phones and tablets). At the same time, the speed of these devices has increased significantly, making it possible to start to process this data in a meaningful manner. Computer Vision is about how we can automate image or video understanding on machines. It covers the techniques used to automate tasks ranging from industrial inspection (where the image understanding problem is constrained to one which we could easily address 20 years ago) to video understanding in order to guide autonomous robots so that they can interact in a meaningful and safe manner in a world designed for humans.

This book provides a brief introduction to this exciting field, covering the basics of image processing and providing the reader with enough information to solve many practical problems. Computer vision systems are becoming ubiquitous. They are in our homes (in the interfaces of the games consoles which our children use), in our cameras and phones (providing automatic face detection and red eye removal), on our streets (determining the licence plates of vehicles passing through toll gates), in our offices (providing biometric verification of identity), and even more so in our factories, helping to guide robots to manufacture goods (such as cars) and automatically inspecting goods to ensure they look right. Yet it seems that we are only at the beginning of how computer vision can be employed, and we can expect significantly more vision systems to emerge.

For those interested in this field as developers (and that hopefully includes you as you are reading this book) there is very good news as there are a number of high quality systems in which computer vision solutions can be developed, of which two stand out in particular: MATLAB® and OpenCV. MATLAB® provides an environment that allows relatively rapid prototyping of vision solutions. OpenCV is a high quality library for C and C++, with wrappers for Python and Java (on Windows, Linux, MacOS, FreeBSD, OpenBSD, Android, Maemo and iOS), which provides implementations of many state-of-the-art vision techniques. OpenCV is the platform of choice for many vision developers, is developed collaboratively by the vision community and is available free of charge for educational and commercial use. OpenCV code snippets are provided throughout this book so that readers can easily take the theory and easily create working solutions to vision problems.

This text is intended to:

1. Provide a solid academic background to basic computer vision.
2. Provide enough material for a one-semester course in computer vision. Larger, all encompassing, textbooks are very off-putting to new students in this (or any) field.
3. Facilitate practical use of computer vision. The goal is to bridge the gap between the theory and the practical implementation of computer vision and so explanations of how to use the relevant OpenCV library routines are included, accompanied by a full working program including the code snippets from the text in the website mentioned below.
4. Allow students to solve real practical problems by providing images and videos for the 20 application problems in Chapter 10.

Electronic Resources

The electronic resources which accompany this text inlcude:

- the code examples from the text along with images generated from the code to give an idea of the processing done by each section of the code.
- Powerpoint slides for each of the chapters.
- the media (images and videos) for each of the application problems in Chapter 10 of the book.
- links to information on OpenCV.

The resources are available at https://www.scss.tcd.ie/publications/book-supplements/A-Practical-Introduction-to-Computer-Vision-with-OpenCV and at (a shorter alternative which redirects to the above page) https://www.scss.tcd.ie/Kenneth.Dawson-Howe/PracticalVisionBook

Teaching Computer Vision Using This Text

A computer vision course based on this text would consist of around 22–28 one-hour lectures together with tutorials and labs. Anticipated lecture hours by topic are as follows:

- Introduction: 1–2 hours
- Images (2.1 Cameras – 2.3 Colour images): 2 hours
- Images (2.4 Noise – 2.5 Smoothing): 2 hours
- Histograms (Chapter 3): 2 hours
- Binary Vision (Chapter 4): 2–3 hours
- Geometric Transformations (Chapter 5): 1–2 hours
- Edges (6.1 Edge detection): 2 hours
- Edges (6.2 Contour segmentation): 1–2 hours
- Edges (6.3 Hough transform): 1–2 hours
- Features (7.1 Moravec corner detection – 7.3 FAST corner detection): 1 hour
- Features (7.4 SIFT): 1 hour
- Recognition (8.1 Template matching and 8.2 Chamfer matching): 1 hour

- Recognition (8.3 Statistical pattern recognition): 1 hour
- Recognition (8.4 Cascade of Haar classifiers): 1 hour
- Recognition (8.6 Performance): 1 hour
- Video (Chapter 9): 2–3 hours

For tutorials, it is suggested that the class be broken into groups of three or four students (all in a single large venue) and that the groups should be asked to come up with solutions to some of the vision problems in Chapter 10 (using the vision techniques they have learnt). The intention is that the students discuss how to solve the problems, coming up with ways of combining the techniques that they have learnt in order to solve them. There is more than one solution to all of the problems, so some of the groups should present their solutions to the class, and the class and lecturer should discuss how appropriate the solutions are. For labs and assignments, the same problems can be used, as OpenCV provides the functionality to allow students to prototype solutions to these problems.

Contents

1

Introduction

Computer vision is the automatic analysis of images and videos by computers in order to gain some understanding of the world. Computer vision is inspired by the capabilities of the human vision system and, when initially addressed in the 1960s and 1970s, it was thought to be a relatively straightforward problem to solve. However, the reason we think/thought that vision is easy is that we have our own visual system which makes the task seem intuitive to our conscious minds. In fact, the human visual system is very complex and even the estimates of how much of the brain is involved with visual processing vary from 25% up to more than 50%.

1.1 A Difficult Problem

The first challenge facing anyone studying this subject is to convince themself that the problem is difficult. To try to illustrate the difficulty, we first show three different versions of the same image in Figure 1.1. For a computer, an image is just an array of values, such as the array shown in the left-hand image in Figure 1.1. For us, using our complex vision system, we can perceive this as a face image but only if we are shown it as a grey scale image (top right).

Computer vision is quite like understanding the array of values shown in Figure 1.1, but is more complicated as the array is really much bigger (e.g. to be equivalent to the human eye a camera would need around 127 million elements), and more complex (i.e. with each point represented by three values in order to encode colour information). To make the task even more convoluted, the images are constantly changing, providing a stream of 50–60 images per second and, of course, there are two streams of data as we have two eyes/cameras.

Another illustration of the difficulty of vision was provided by psychologist John Wilding considering his own visual system:

> *As I look out of my window, I see grass and trees, gently swaying in the wind, with a lake beyond . . . An asphalt path leads down through the trees to the lake and two squirrels are chasing each other to and fro across it, ignoring the woman coming up the path . . .*

A Practical Introduction to Computer Vision with OpenCV, First Edition. Kenneth Dawson-Howe.

67	67	66	68	66	67	64	65	65	63	63	69	61	64	63	66	61	60
69	68	63	68	65	62	65	61	50	26	32	65	61	67	64	65	66	63
72	71	70	87	67	60	28	21	17	18	13	15	20	59	61	65	66	64
75	73	76	78	67	26	20	19	16	18	16	13	18	21	50	61	69	70
74	75	78	74	39	31	31	30	46	37	69	66	64	43	18	63	69	60
73	75	77	64	41	20	18	22	63	92	99	88	78	73	39	40	59	65
74	75	71	42	19	12	14	28	79	102	107	96	87	79	57	29	68	66
75	75	66	43	12	11	16	62	87	84	84	108	83	84	59	39	70	66
76	74	49	42	37	10	34	78	90	99	68	94	97	51	40	69	72	65
76	63	40	57	123	88	60	83	95	88	80	71	67	69	32	67	73	73
78	50	32	33	90	121	66	86	100	116	87	85	80	74	71	56	58	48
80	40	33	16	63	107	57	86	103	113	113	104	94	86	77	48	47	45
88	41	35	10	15	94	67	96	98	91	86	105	81	77	71	35	45	47
87	51	35	15	15	17	51	92	104	101	72	74	87	100	27	31	44	46
86	42	47	11	13	16	71	76	89	95	116	91	67	87	12	25	43	51
96	67	20	12	17	17	86	89	90	101	96	89	62	13	11	19	40	51
99	88	19	15	15	18	32	107	99	86	95	92	26	13	13	16	49	52
99	77	16	14	14	16	35	115	111	109	91	79	17	16	13	46	48	51

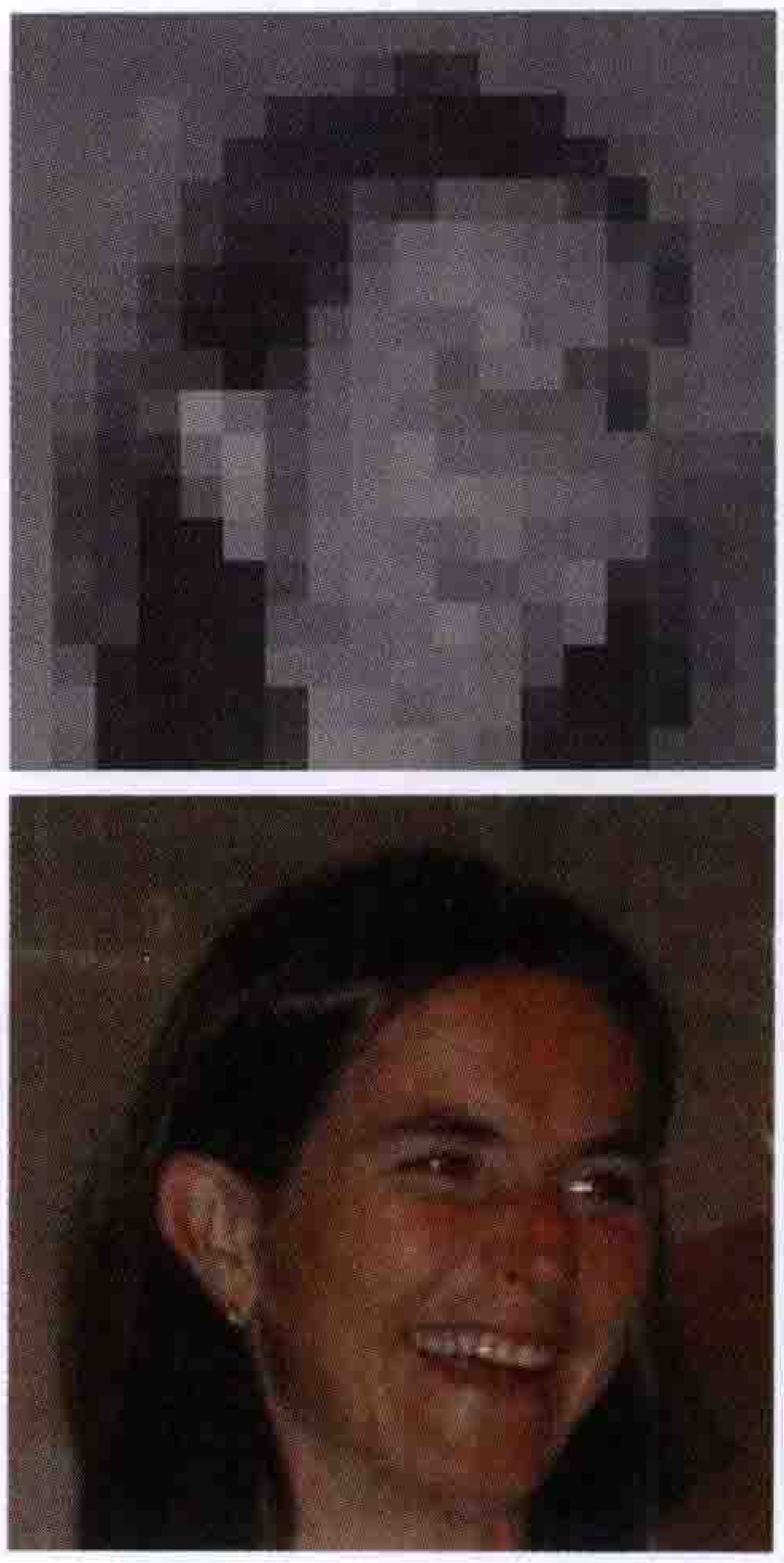


Figure 1.1 Different versions of an image. An array of numbers (left) which are the values of the grey scales in the low resolution image of a face (top right). The task of computer vision is most like understanding the array of numbers

> *This is the scene I experience, a world of objects with background, acted upon and sometimes acting and interacting in events. I have no problem seeing and hearing and smelling and feeling all these things because they affect my senses directly and they make up the real world.*
>
> *Or do they? I can look again and notice things I missed before, or see the scene in new ways. There is a white wall framing the window I am looking through and the window in fact fills less of my field of view that the wall, but I did not even notice the wall at first, and my impression was that the scene through the window was a panorama right across in front of me. There are metal bars dividing the window into squares and the glass is obscured with dust and spots but for me the view seems complete and un-obscured. The 'grass' is patches of colour ranging from nearly white in the bright sun to nearly black in the shade but I 'saw' green grass in light and shade. Other changing greenish shapes were for me permanent leafy branches moved by a wind I neither saw nor felt, and two constantly varying grey shapes were squirrels moving with a purpose. Another shape increasing in size and changing in position was an approaching woman.* (Wilding, 1983)

1.2 The Human Vision System

If we could duplicate the human visual system then the problem of developing a computer vision system would be solved. So why can't we? The main difficulty is that we do not understand what the human vision system is doing most of the time.

If you consider your eyes, it is probably not clear to you that your colour vision (provided by the 6–7 million cones in the eye) is concentrated in the centre of the visual field of the eye (known as the macula). The rest of your retina is made up of around 120 million rods (cells that are sensitive to visible light of any wavelength/colour). In addition, each eye has a rather large blind spot where the optic nerve attaches to the retina. Somehow, we think we see a continuous image (i.e. no blind spot) with colour everywhere, but even at this lowest level of processing it is unclear as to how this impression occurs within the brain.

The visual cortex (at the back of the brain) has been studied and found to contain cells that perform a type of edge detection (see Chapter 6), but mostly we know what sections of the brain do based on localised brain damage to individuals. For example, a number of people with damage to a particular section of the brain can no longer recognise faces (a condition known as prosopagnosia). Other people have lost the ability to sense moving objects (a condition known as akinetopsia). These conditions inspire us to develop separate modules to recognise faces (e.g. see Section 8.4) and to detect object motion (e.g. see Chapter 9).

We can also look at the brain using functional MRI, which allows us to see the concentration of electrical activity in different parts of the brain as subjects perform various activities. Again, this may tell us what large parts of the brain are doing, but it cannot provide us with algorithms to solve the problem of interpreting the massive arrays of numbers that video cameras provide.

1.3 Practical Applications of Computer Vision

Computer vision has many applications in industry, particularly allowing the automatic inspection of manufactured goods at any stage in the production line. For example, it has been used to:

- Inspect printed circuits boards to ensure that tracks and components are placed correctly. See Figure 1.2.
- Inspect print quality of labels. See Figure 1.3.
- Inspect bottles to ensure they are properly filled. See Figure 1.3.

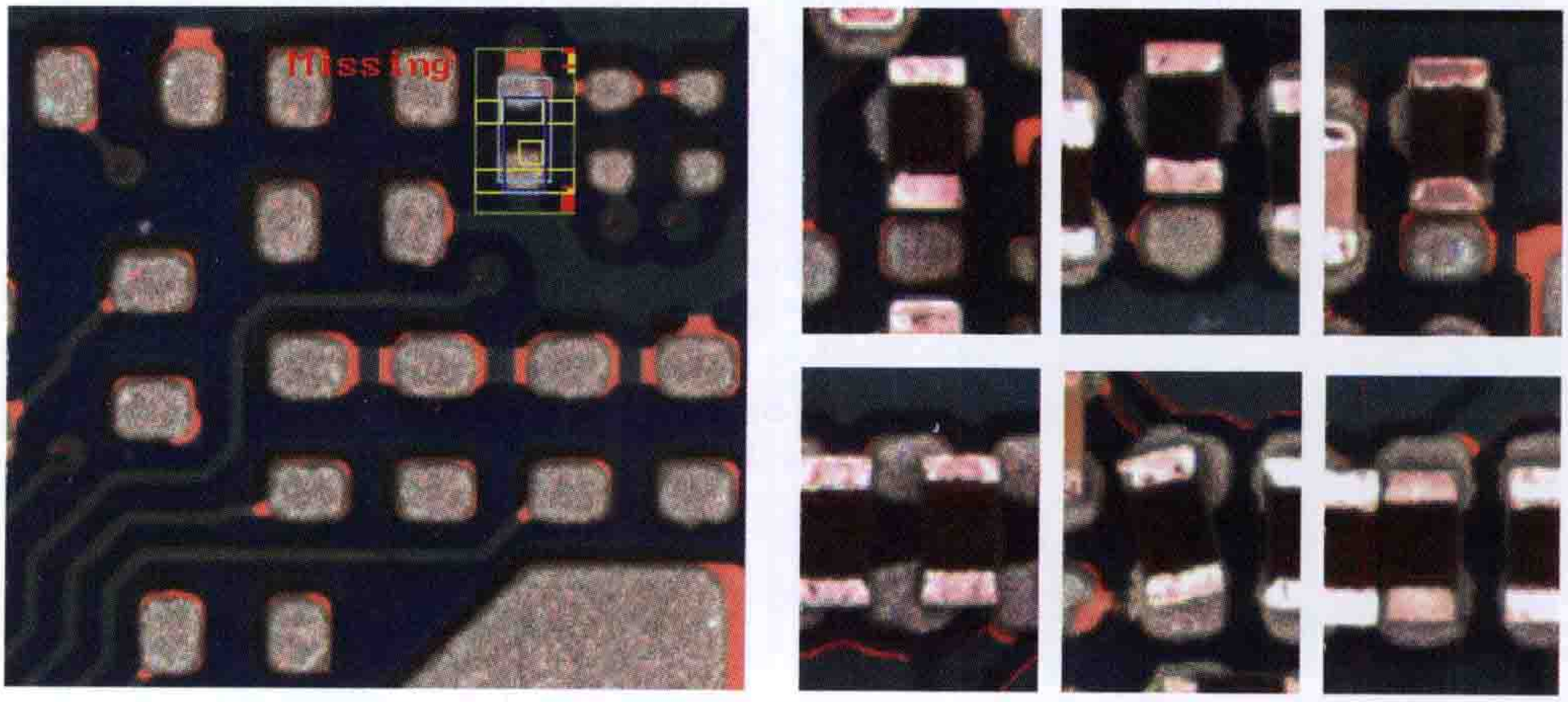


Figure 1.2 PCB inspection of pads (left) and images of some detected flaws in the surface mounting of components (right). Reproduced by permission of James Mahon

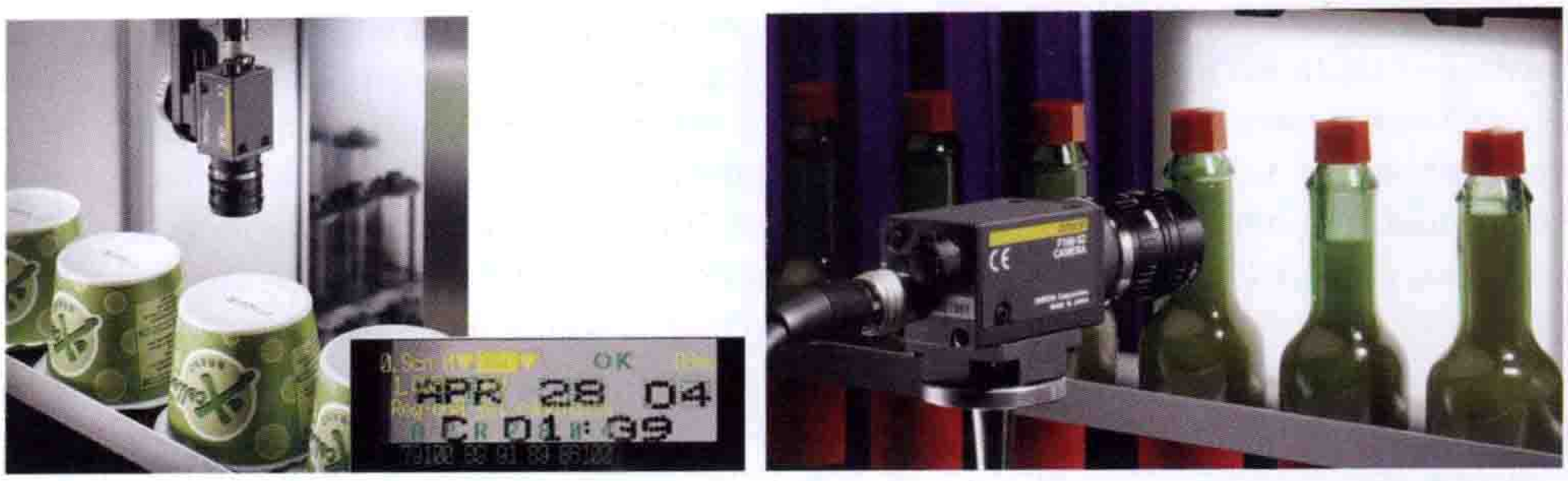


Figure 1.3 Checking print quality of best-before dates (right), and monitoring level to which bottles are filled (right). Reproduced by permission of Omron Electronics LLC

- Inspect apples to determine if there is any bruising.
- Locate chocolates on a production line so that a robot arm can pick them up and place them in the correct locations in the box.
- Guide robots when manufacturing complex products such as cars.

On the factory floor, the problem is a little simpler than in the real world as the lighting can be constrained and the possible variations of what we can see are quite limited. Computer vision is now solving problems outside the factory. Computer vision applications outside the factory include:

- The automatic reading of license plates as they pass through tollgates on major roads.
- Augmenting sports broadcasts by determining distances for penalties, along with a range of other statistics (such as how far each player has travelled during the game).
- Biometric security checks in airports using images of faces and images of fingerprints. See Figure 1.4.
- Augmenting movies by the insertion of virtual objects into video sequences, so that they appear as though they belong (e.g. the candles in the Great Hall in the Harry Potter movies).

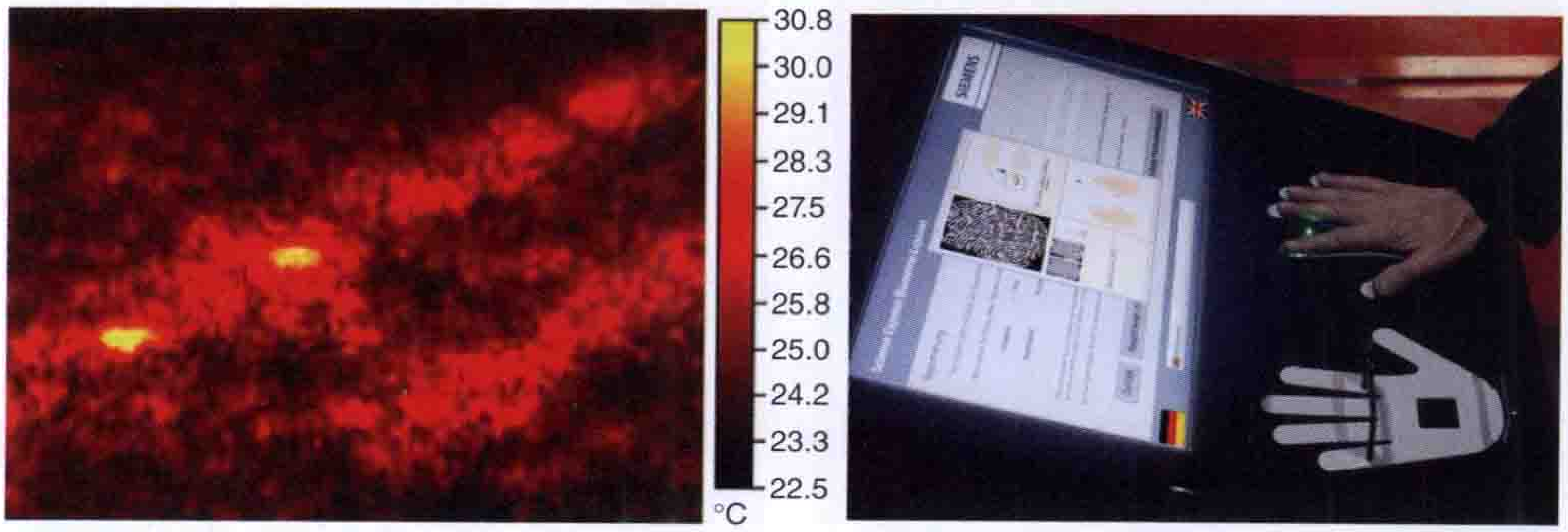


Figure 1.4 Buried landmines in an infrared image (left). Reproduced by permission of Zouheir Fawaz, Handprint recognition system (right). Reproduced by permission of Siemens AG