# 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 |

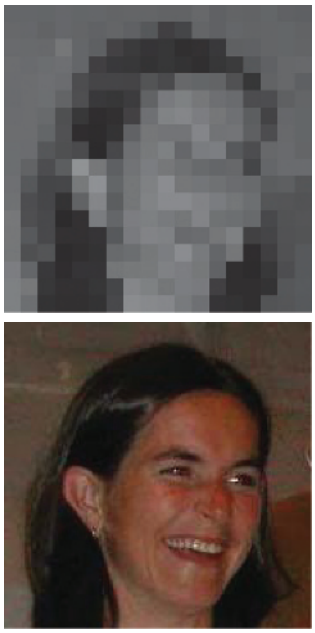

**Figure 1.1** Different versions of an image. An array of numbers (left) which are the values of the grey scales in the low resolution image of a face (top right). The task of computer vision is most like understanding the array of numbers

*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.

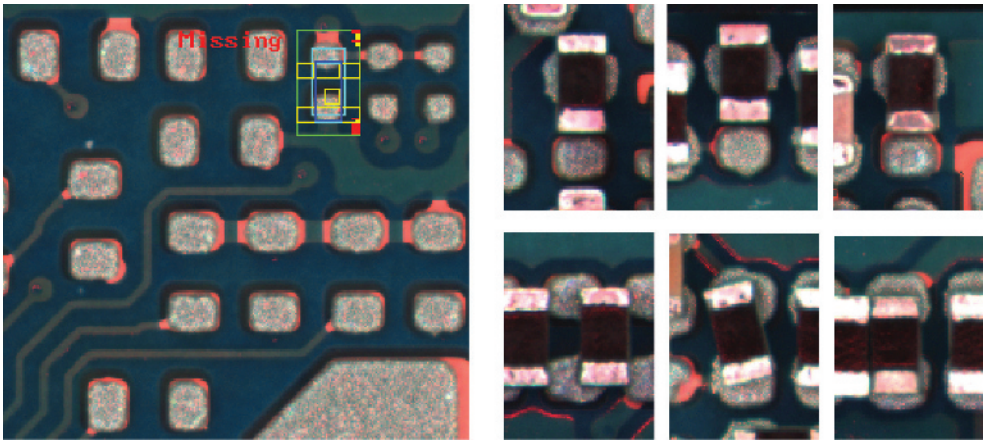


**Figure 1.2** PCB inspection of pads (left) and images of some detected flaws in the surface mounting of components (right). Reproduced by permission of James Mahon


**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).

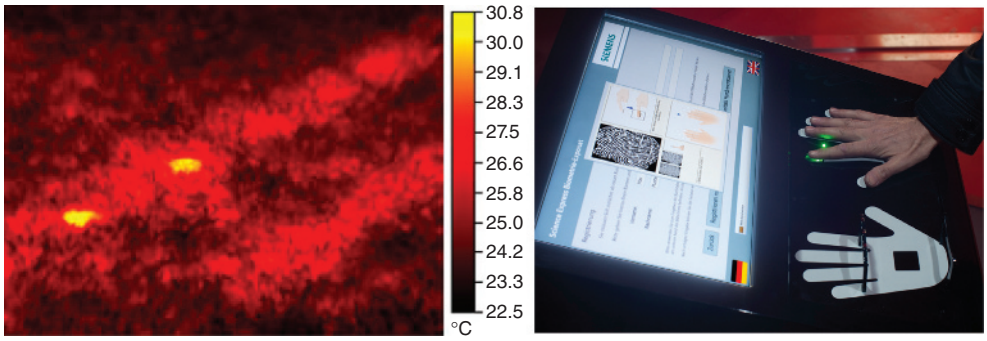


**Figure 1.4** Buried landmines in an infrared image (left). Reproduced by permission of Zouheir Fawaz, Handprint recognition system (right). Reproduced by permission of Siemens AG

- Assisting drivers by warning them when they are drifting out of lane.
- Creating 3D models of a destroyed building from multiple old photographs.
- Advanced interfaces for computer games allowing the real time detection of players or their hand-held controllers.
- Classification of plant types and anticipated yields based on multispectral satellite images.
- Detecting buried landmines in infrared images. See Figure 1.4.

Some examples of existing computer vision systems in the outside world are shown in Figure 1.4.

## 1.4 The Future of Computer Vision

The community of vision developers is constantly pushing the boundaries of what we can achieve. While we can produce autonomous vehicles, which drive themselves on a highway, we would have difficulties producing a reliable vehicle to work on minor roads, particularly if the road marking were poor. Even in the highway environment, though, we have a legal issue, as who is to blame if the vehicle crashes? Clearly, those developing the technology do not think it should be them and would rather that the driver should still be responsible should anything go wrong. This issue of liability is a difficult one and arises with many vision applications in the real world. Taking another example, if we develop a medical imaging system to diagnose cancer, what will happen when it mistakenly does not diagnose a condition? Even though the system might be more reliable than any individual radiologist, we enter a legal minefield. Therefore, for now, the simplest solution is either to address only non-critical problems or to develop systems, which are assistants to, rather than replacements for, the current human experts.

Another problem exists with the deployment of computer vision systems. In some countries the installation and use of video cameras is considered an infringement of our basic right to privacy. This varies hugely from country to country, from company to company, and even from individual to individual. While most people involved with technology see the potential benefits of camera systems, many people are inherently distrustful of video cameras and what the videos *could* be used for. Among other things, they fear (perhaps justifiably) a Big Brother scenario, where our movements and actions are constantly monitored. Despite this, the number of cameras is growing very rapidly, as there are cameras on virtually every new computer, every new phone, every new games console, and so on.

Moving forwards, we expect to see computer vision addressing progressively harder problems; that is problems in more complex environments with fewer constraints. We expect computer vision to start to be able to recognise more objects of different types and to begin to extract more reliable and robust descriptions of the world in which they operate. For example, we expect computer vision to

- become an integral part of general computer interfaces;
- provide increased levels of security through biometric analysis;
- provide reliable diagnoses of medical conditions from medical images and medical records;
- allow vehicles to be driven autonomously;
- automatically determine the identity of criminals through the forensic analysis of video.

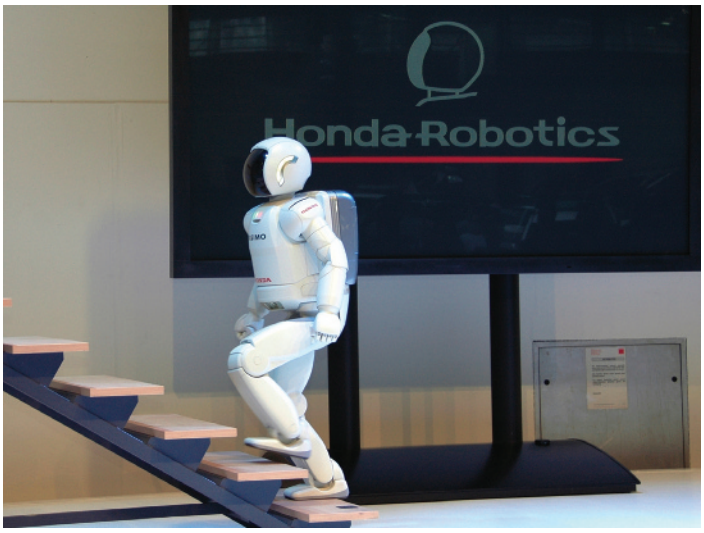


**Figure 1.5** The ASIMO humanoid robot which has two cameras in its 'head' which allow ASIMO to determine how far away things are, recognise familiar faces, etc. Reproduced by permission of Honda Motor Co. Inc

Ultimately, computer vision is aiming to emulate the capabilities of human vision, and to provide these abilities to humanoid (and other) robotic devices, such as ASIMO (see Figure 1.5). This is part of what makes this field exciting, and surprising, as we all have our own (human) vision systems which work remarkably well, yet when we try to automate any computer vision task it proves very difficult to do reliably.

## 1.5 Material in this Textbook

This textbook is intended to provide an illustrated introduction to the area of computer vision. It provides roughly the amount of material which can be covered in a one-semester, year four or five university course. While this text covers the theory behind basic computer vision, it also provides a bridge from the theory to practical implementation using the industry standard OpenCV libraries (by explaining how the operations can be invoked in OpenCV).

In Chapter 2, we consider the basics of cameras and images, along with consideration of the noise that is exhibited by many images and the techniques through which this noise can be removed or attenuated.

In Chapter 3, we consider how image information can be summarised in the form of a histogram and how those histograms can be used in order to enhance images or to extract information from the images.

Chapter 4 looks at the most commonly used technique for industrial vision – that of binary vision, where we simplify images so that every point is either black or white. This approach makes the processing much easier (assuming that the binary image can be obtained correctly).

Chapter 5 looks at how we model and remove distortion from images and cameras (which is often introduced by the camera/lens system).

Chapter 6 describes the extraction and use of edges (locations at which the brightness or colour changes significantly) in images. These cartoon-like features allow us to abstract information from images. Edge detection does not perform well at corners, and in Chapter 7 we look at corner/feature points that can act as a complement to edges or can be used on their own to provide less ambiguous features with which to match different images or objects.

In Chapter 8, we look at a number of common approaches to recognition in images, as in many applications we need to determine the location and identity of objects (e.g. license plates, faces, etc.).

Chapter 9 looks at the basics of processing videos, concentrating particularly on how we detect moving objects in video feeds from static cameras (a problem that occurs frequently in video surveillance), how we track objects from frame to frame and how we can assess performance in video processing.

Finally, in Chapter 10, we present a large number of vision application problems to provide students with the opportunity to solve real problems (which is the only way to really appreciate how difficult computer vision is). Images or videos for these problems are provided in the resources associated with this book.

We end this introduction with a quote from the 19th century: '… apprehension by the senses supplies after all, directly or indirectly, the material of all human knowledge, or at least the stimulus necessary to develop every inborn faculty of the mind. It supplies the basis for the whole action of man upon the outer world … For there is little hope that he who does not begin at the beginning of knowledge will ever arrive at its end' (von Helmholtz, 1868).

## 1.6 Going Further with Computer Vision

This text contains roughly the amount of material that could be covered in a one semester introductory course on computer vision. There are many very large comprehensive vision texts, which allow students to delve more deeply and broadly into computer vision. A few that come highly recommended are (Sonka, et al., 2007), (González & Woods, 2007), and (Marques, 2011).

For those wishing to go further with computer vision and OpenCV, I would recommend the practical textbook by Baggio (Daniel Lélis Baggio, 2012). This leads the reader through a series of relatively advanced vision application problems on a variety of platforms.