Part IV Computer Vision

Part

IV Computer Vision

Vision is the process of discovering from images
what is present in the world and where it is.
David Marr

Fig. IV.1. a Robber fly, *Holocephala fusca*; **b** jumping spider, *Phidippus putnami* (**a** and **b** courtesy Thomas Shahan, thomasshanan.com). **c** Scallop (courtesy Sönke Johnsen), each of the small blue spheres is an eye. **d** Human eye

Almost all animal species use eyes – in fact evolution has *invented* the eye many times over. Figure IV.1 shows a variety of eyes from nature: the compound eye of a fly, the main and secondary eyes of a spider, the reflector-based eyes of a scallop, and the lens-based eye of a human. Vertebrates have two eyes, but spiders and scallops have many eyes.

Even very simple animals, bees for example, with brains comprising just 10^6 neurons (compared to our 10^{11}) are able to perform complex and life critical tasks such as finding food and returning it to the hive using vision (Srinivasan and Venkatesh 1997). This is despite the very high biological cost of owning an eye: the complex eye itself, muscles to move it, eyelids and tear ducts to protect it, and a large visual cortex (relative to body size) to process its data.

Our own experience is that eyes are very effective sensors for recognition, navigation, obstacle avoidance and manipulation. Cameras mimic the function of an eye and we wish to use cameras to create vision-based competencies for robots – to use digital images to recognize objects and navigate within the world. Figure IV.2 shows a robot with a number of different types of cameras.

Technological development has made it feasible for robots to use cameras as eyes. For much of the history of computer vision, dating back to the 1960s, electronic cameras were cumbersome and expensive and computer power was inadequate. Today CMOS cameras for cell phones cost just a few dollars each, and our mobile and personal computers come standard with massive parallel computing power. New algorithms, cheap sensors and plentiful computing power make vision a practical sensor today.

In Chap. 1 we defined a robot as

a goal oriented machine that can sense, plan and act

and this part of the book is concerned with sensing using vision, or visual perception. Whether a robot works in a factory or a field it needs to sense its world in order to plan its actions.

In this part of the book we will discuss the process of vision from start to finish: from the light falling on a scene, being reflected, gathered by a lens, turned into a digital image and processed by various algorithms to extract the information required to support the robot competencies listed above. These steps are depicted graphically in Fig. IV.3.

Development of the eye. It is believed that all animal eyes share a common ancestor in a proto-eye that evolved 540 million years ago. However major evolutionary advances seem to have occurred in just the last few million years. The very earliest eyes, called eyespots, were simple patches of photoreceptor protein in single-celled animals. Multi-celled animals evolved multi-cellular eyespots which could sense the brightness of light but not its direction. Gradually the eyespot evolved into a shallow cup shape which gave a limited ability to discriminate directional brightness according to which cells were illuminated. The pit deepened, the opening became smaller, and the number of photoreceptor cells increased, forming a pin-hole camera that was capable of distinguishing shapes. Next came an overgrowth of transparent cells to protect the eyespot which led to a filled eye chamber and eventually the eye as we know it today. The lensed eye has evolved independently seven different times across species. Nature has evolved ten quite distinct eye designs including those shown above.

Fig. IV.2.
A cluster of cameras on an outdoor mobile robot: forward looking stereo pair, side looking wide angle camera, overhead panoramic camera mirror (CSIRO mobile robot)

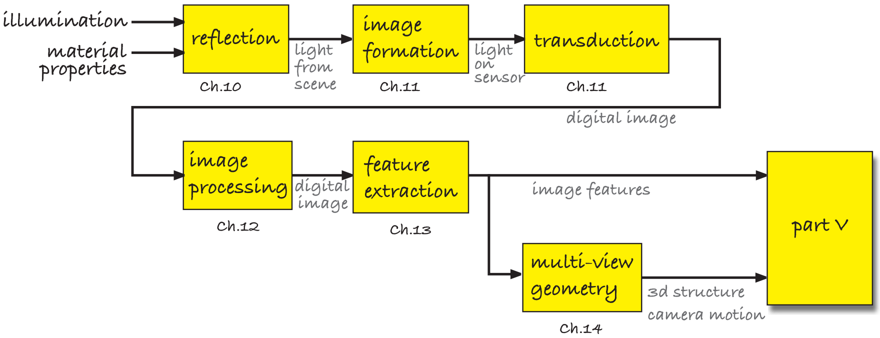

Fig. IV.3.
Steps involved in image processing

In Chap. 10 we start by discussing light, and in particular color because it is such an important characteristic of the world that we perceive. Although we learn about color at kindergarten it is a complex topic that is often not well understood. Next, in Chap. 11, we discuss how an image of the world is formed on a sensor and converted to a digital image that can be processed by a computer. Fundamental image processing algorithms are covered in Chap. 12 and provide the foundation for the feature extraction algorithms discussed in Chap. 13. Feature extraction is a problem in data reduction, in extracting the *essence* of the scene from the massive amount of pixel data. For example, how do we determine the coordinate of the round red object in the scene, which can be described with perhaps just 4 bytes, given the millions of bytes that comprise an image. To solve this we must address many important subproblems such as "what is red?", "how do we distinguish red pixels from nonred pixels?", "how do we describe the shape of the red pixels?", "what if there are more than one red object?" and so on.

As we progress through these chapters we will encounter the limitations of using just a single camera to view the world. Once again biology shows the way – multiple eyes are common and have great utility. This leads us to consider using multiple views of the world, from a single moving camera or multiple cameras observing the scene from different viewpoints. This is discussed in Chap. 14 and is particularly important for understanding the 3-dimensional structure of the world. All of this sets the scene for describing how vision can be used for closed-loop control of arm-type and mobile robots which is the subject of the next and final part of the book.