

Part I

COPYRIGHTED MATERIAL

1

Introduction

The purpose of this text on stereo-based imaging is twofold: it is to give students of computer vision a thorough grounding in the image analysis and projective geometry techniques relevant to the task of recovering three-dimensional (3D) surfaces from stereo-pair images; and to provide a complete reference text for professional researchers in the field of computer vision that encompasses the fundamental mathematics and algorithms that have been applied and developed to allow 3D vision systems to be constructed.

Prior to reviewing the contents of this text, we shall set the context of this book in terms of the underlying objectives and the explanation and design of 3D vision systems. We shall also consider briefly the historical context of optics and vision research that has led to our contemporary understanding of 3D vision.

Here we are specifically considering 3D vision systems that base their operation on acquiring stereo-pair images of a scene and then decoding the depth information *implicitly* captured within the stereo-pair as parallaxes, i.e. relative displacements of the contents of one of the images of the stereo-pair with respect to the other image. This process is termed *stereo-photogrammetry*, i.e. measurement from stereo-pair images. For readers with normal functional binocular vision, the everyday experience of observing the world with both of our eyes results in the perception of the relative distance (depth) to points on the surfaces of objects that enter our field of view. For over a hundred years it has been possible to configure a stereo-pair of cameras to capture stereo-pair images, in a manner analogous to mammalian binocular vision, and thereafter view the developed photographs to observe a miniature 3D scene by means of a stereoscope device (used to present the left and right images of the captured stereo-pair of photographs to the appropriate eye). However, in this scenario it is the brain of the observer that must decode the depth information locked within the stereo-pair and thereby experience the perception of depth. In contrast, in this book we shall present underlying mechanisms by which a computer program can be devised to analyse digitally formatted images captured by a stereo-pair of cameras and thereby recover an *explicit* measurement of distances to points *sampling* surfaces in the imaged field of view. Only by explicitly recovering depth estimates does it become possible to undertake useful tasks such as 3D measurement or reverse engineering of object surfaces as elaborated below. While the science of stereo-photogrammetry is a well-established field and it has indeed been possible to undertake 3D

measurement by means of stereo-pair images using a manually operated measurement device (the stereo-comparator) since the beginning of the twentieth century, we present fully automatic approaches for 3D imaging and measurement in this text.

1.1 Stereo-pair Images and Depth Perception

To appreciate the structure of 3D vision systems based on processing stereo-pair images, it is first necessary to grasp, at least in outline, the most basic principles involved in the formation of stereo-pair images and their subsequent analysis. As outlined above, when we observe a scene with both eyes, an image of the scene is formed on the retina of each eye. However, since our eyes are horizontally displaced with respect to each other, the images thus formed are not identical. In fact this stereo-pair of retinal images contains slight displacements between the relative locations of local parts of the image of the scene with respect to each image of the pair, depending upon how close these local scene components are to the point of *fixation* of the observer's eyes. Accordingly, it is possible to reverse this process and deduce how far away scene components were from the observer according to the magnitude and direction of the parallaxes within the stereo-pairs when they were captured. In order to accomplish this task two things must be determined: firstly, those local parts of one image of the stereo-pair that match the corresponding parts in the other image of the stereo-pair, in order to find the local parallaxes; secondly, the precise geometric properties and configuration of the eyes, or cameras. Accordingly, a process of *calibration* is required to discover the requisite geometric information to allow the imaging process to be inverted and relative distances to surfaces observed in the stereo-pair to be recovered.

1.2 3D Vision Systems

By definition, a stereo-photogrammetry-based 3D vision system will require stereo-pair image acquisition hardware, usually connected to a computer hosting software that automates acquisition control. Multiple stereo-pairs of cameras might be employed to allow all-round coverage of an object or person, e.g. in the context of whole-body scanners. Alternatively, the object to be imaged could be mounted on a computer-controlled turntable and overlapping stereo-pairs captured from a fixed viewpoint for different turntable positions. Accordingly, sequencing capture and image download from multiple cameras can be a complex process, and hence the need for a computer to automate this process.

The stereo-pair acquisition process falls into two categories, active illumination and passive illumination. Active illumination implies that some form of pattern is projected on to the scene to facilitate finding and disambiguating parallaxes (also termed *correspondences* or *disparities*) between the stereo-pair images. Projected patterns often comprise grids or stripes and sometimes these are even colour coded. In an alternative approach, a random speckle texture pattern is projected on to the scene in order to augment the texture already present on imaged surfaces. Speckle projection can also guarantee that that imaged surfaces appear to be randomly textured and are therefore locally uniquely distinguishable and hence able to be matched successfully using certain classes of image matching algorithm. With the advent of 'high-resolution' digital cameras the need for pattern projection has been reduced, since the surface texture naturally present on materials, having even a matte finish, can serve to facilitate

matching stereo-pairs. For example, stereo-pair images of the human face and body can be matched successfully using ordinary studio flash illumination when the pixel sampling density is sufficient to resolve the natural texture of the skin, e.g. skin-pores. A camera resolution of approximately 8–13M pixels is adequate for stereo-pair capture of an area corresponding to the adult face or half-torso.

The acquisition computer may also host the principal 3D vision software components:

- An image matching algorithm to find correspondences between the stereo-pairs.
- Photogrammetry software that will perform system calibration to recover the geometric configuration of the acquisition cameras and perform 3D point reconstruction in world coordinates.
- 3D surface reconstruction software that builds complete manifolds from 3D *point-clouds* captured by each imaging stereo-pair.

3D visualisation facilities are usually also provided to allow the reconstructed surfaces to be displayed, often *draped* with an image to provide a *photorealistic* surface model. At this stage the 3D shape and surface appearance of the imaged object or scene has been captured in explicit digital metric form, ready to feed some subsequent application as described below.

1.3 3D Vision Applications

This book has been motivated in part by the need to provide a manual of techniques to serve the needs of the computer vision practitioner who wishes to construct 3D imaging systems configured to meet the needs of practical applications. A wide variety of applications are now emerging which rely on the fast, efficient and low-cost capture of 3D surface information. The traditional role for image-based 3D surface measurement has been the reserve of *close-range* photogrammetry systems, capable of recovering surface measurements from objects in the range of a few tens of millimetres to a few metres in size. A typical example of a classical close-range photogrammetry task might comprise surface measurement for manufacturing quality control, applied to high-precision engineered products such as aircraft wings.

Close-range video-based photogrammetry, having a lower spatial resolution than traditional plate-camera film-based systems, initially found a niche in imaging the human face and body for clinical and creative media applications. 3D clinical photographs have the potential to provide quantitative measurements that reduce subjectivity in assessing the surface anatomy of a patient (or animal) before and after surgical intervention by providing numeric, possibly automated, scores for the shape, symmetry and longitudinal change of anatomic structures. Creative media applications include whole-body 3D imaging to support creation of human avatars of specific individuals, for 3D gaming and cine special effects requiring virtual actors. Clothing applications include body or foot scanning for the production of custom clothing and shoes or as a means of sizing customers accurately. An innovative commercial application comprises a ‘virtual catwalk’ to allow customers to visualize themselves in clothing prior to purchasing such goods on-line via the Internet.

There are very many more emerging uses for 3D imaging beyond the above and commercial ‘reverse engineering’ of premanufactured goods. 3D vision systems have the potential to revolutionize autonomous vehicles and the capabilities of robot vision systems. Stereo-pair cameras could be mounted on a vehicle to facilitate autonomous navigation or configured

within a robot workcell to endow a ‘blind’ pick-and-place robot, both object recognition capabilities based on 3D cues and simultaneously 3D spatial quantification of object locations in the workspace.

1.4 Contents Overview: The 3D Vision Task in Stages

The organization of this book reflects the underlying principles, structural components and uses of 3D vision systems as outlined above, starting with a brief historical view of vision research in Chapter 2. We deal with the basic existence proof that binocular 3D vision is possible, in an overview of the human visual system in Chapter 3. The basic projective geometry techniques that underpin 3D vision systems are also covered here, including the geometry of monocular and binocular image formation which relates how binocular parallaxes are produced in stereo-pair images as a result of imaging scenes containing variation in depth. Camera calibration techniques are also presented in Chapter 3, completing the introduction of the role of image formation and geometry in the context of 3D vision systems.

We deal with fundamental 2D image analysis techniques required to undertake image filtering and feature detection and localization in Chapter 4. These topics serve as a precursor to perform image matching, the process of detecting and quantifying parallaxes between stereo-pair images, a prerequisite to recovering depth information. In Chapter 5 the issue of spatial scale in images is explored, namely how to structure algorithms capable of efficiently processing images containing structures of varying scales which are unknown in advance. Here the concept of an image *scale-space* and the *multi-resolution image pyramid* data structure is presented, analysed and explored as a precursor to developing matching algorithms capable of operating over a wide range of visual scales. The core algorithmic issues associated with stereo-pair image matching are contained in Chapter 6 dealing with distance measures for comparing image patches, the associated parametric issues for matching and an in-depth analysis of area-based matching over scale-space within a practical matching algorithm. Feature-based approaches to matching are also considered and their combination with area-based approaches. Then two solutions to the stereo problem are discussed: the first, based on the *dynamic programming*, and the second one based on the *graph cuts* method. The chapter ends with discussion of the *optical flow* methods which allow estimation of local displacements in a sequence of images.

Having dealt with the recovery of disparities between stereo-pairs, we progress logically to the recovery of 3D surface information in Chapter 7. We consider the process of *triangulation* whereby 3D points in world coordinates are computed from the disparities recovered in the previous chapter. These 3D points can then be organized into surfaces represented by *polygonal meshes* and the *3D point-clouds* recovered from *multi-view* systems acquiring more than one stereo-pair of the scene can be fused into a coherent surface model either directly or via volumetric techniques such as *marching cubes*. In Chapter 8 we conclude the progression from theory to practice, with a number of case examples of 3D vision applications covering areas such as face and body imaging for clinical, veterinary and creative media applications and also 3D vision as a visual prosthetic. An application based only on image matching is also presented that utilizes motion-induced inter-frame disparities within a cine sequence to synthesize missing or damaged frames, or sets of frames, in digitized historic archive footage.



Figure 1.1 depicts possible order of reading the book. All chapters can be read in number order or selectively as references to specific topics. There are five main chapters (Chapters 3–7), three auxiliary chapters (Chapters 1, 2 and 8) as well as five technical tutorials (Chapters 9–13). The latter are intended to aid understanding of specific topics and can be read in conjunction with the related main chapters, as indicated by the dashed lines in Figure 1.1.

