1

Chapter

Introduction

1.1 MOTIVATION

Vision allows humans to perceive and understand the world surrounding them, while computer vision aims to duplicate the effect of human vision by electronically perceiving and understanding an image. Books other than this one would dwell at length on this sentence and the meaning of the word 'duplicate'—whether computer vision is *simulating* or *mimicking* human systems is philosophical territory, and very fertile territory too.

Giving computers the ability to see is not an easy task—we live in a three-dimensional (3D) world, and when computers try to analyze objects in 3D space, the visual sensors available (e.g., TV cameras) usually give two-dimensional (2D) images, and this projection to a lower number of dimensions incurs an enormous loss of information. Sometimes, equipment will deliver images that are 3D but this may be of questionable value: analyzing such datasets is clearly more complicated than 2D, and sometimes the 'three-dimensionality' is less than intuitive to us ... terahertz scans are an example of this. Dynamic scenes such as those to which we are accustomed, with moving objects or a moving camera, are increasingly common and represent another way of making computer vision more complicated.

Figure 1.1 could be witnessed in thousands of farmyards in many countries, and serves to illustrate just some of the problems that we will face.

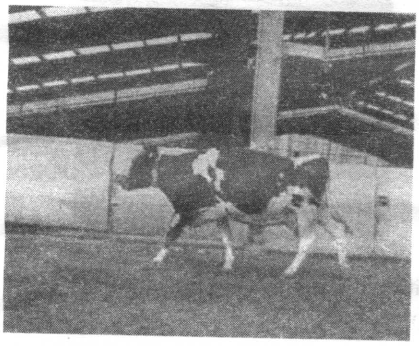

Figure 1.1: A frame from a video of a typical farmyard scene: the cow is one of a number walking naturally from right to left. *Courtesy of D. R. Magee, University of Leeds.*

There are many reasons why we might wish to study scenes such as this, which are attractively simple *to us*—the beast is moving slowly, it is clearly black and white, its movement is rhythmic, ... However, automated analysis is very fraught; in fact, the animal's boundary is often very difficult to distinguish clearly from the background, the motion of the legs is self occluding and (subtly) the concept of 'cow-shaped' is not something easily encoded. The application from which this picture was taken[1] made use of many of the algorithms presented in this book: starting at a low level, moving features were identified and grouped. A 'training phase' taught the system what a cow might look like in various poses (see Figure 1.2), from which a model of a 'moving' cow could be derived (see Figure 1.3).

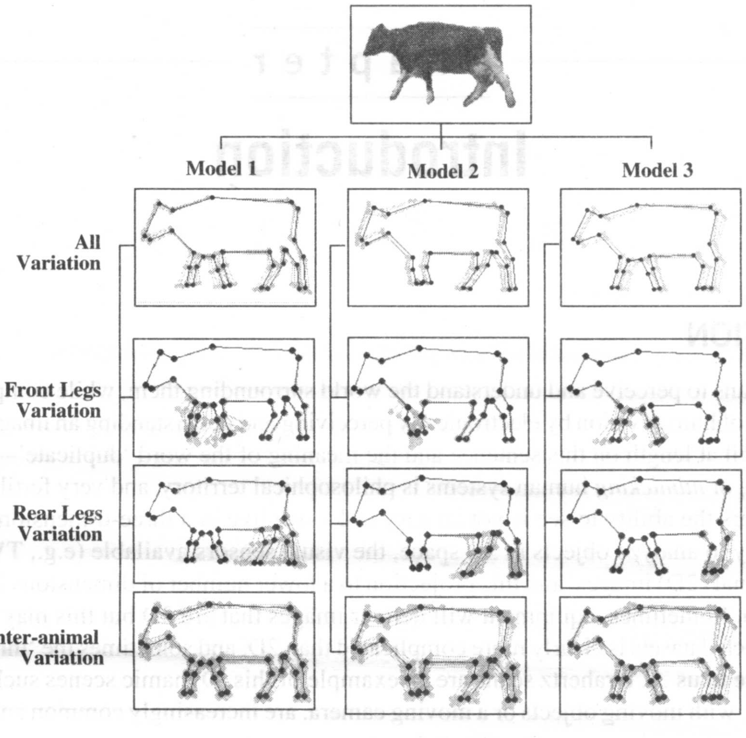

Figure 1.2: Various models for a cow silhouette: a straight-line boundary approximation has been learned from training data and is able to adapt to different animals and different forms of occlusion. *Courtesy of D. R. Magee, University of Leeds.*

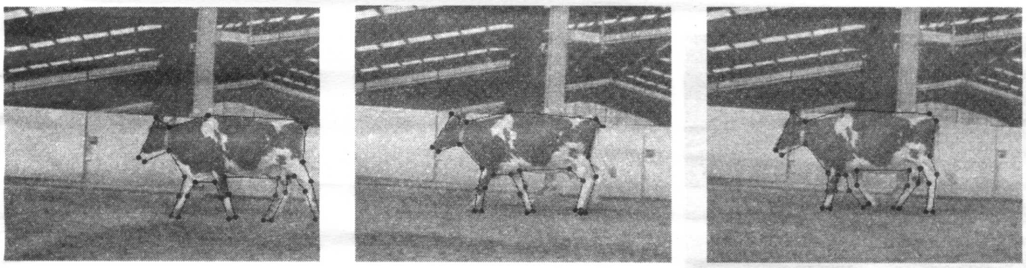

Figure 1.3: Three frames from a cow sequence: notice the model can cope with partial occlusion as the animal enters the scene, and the different poses exhibited. *Courtesy of D. R. Magee, University of Leeds.*

These models could then be fitted to new ('unseen') video sequences. Crudely, at this stage anomalous behavior such as lameness could be detected by the model failing to fit properly, or well.

[1] The application was serious; there is a growing need in modern agriculture for automatic monitoring of animal health, for example to spot lameness. A limping cow is trivial for a human to identify, but it is very challenging to do this automatically [Magee, 2001].