Robot Science

Andrew Davison

January 2008

Contents

Chapter 1

Intelligent Machines

It is a very important time in the field of robotics. New theories and techniques in combination with the ever-increasing performance of modern processors are at last giving robots and other artificial devices the power to interact automatically and usefully with their real, complex surroundings. Today's research robots no longer aim just to execute sequences of car-spraying commands or find their way out of contrived mazes, but to move independently through the world and interact with it in human-like ways. Artificially intelligent devices are now starting to emerge from laboratories and apply their abilities to real-world tasks, and it seems inevitable to me that this will continue at a faster and faster rate until they become ubiquitous everyday objects.

The effects that *artificial intelligence* (AI), robotics and other technologies developing in parallel will have on human society, in the relatively near future, are much greater than most people imagine — for better or worse. Ray Kurzweil, in his book 'The Singularity is Near', presents a strong case that the progress of technology from ancient times to the current day is following an *exponential* trend. An exponential curve in mathematics is one whose height increases by a constant multiplicative factor for each unit horizontal step. The implication is that technological progress gets steadily faster and faster every year. A well-known modern instantiation of this phenomenon is 'Moore's Law', suggested following the observation in 1965 by Gordon Moore, co-founder of the processor company Intel, that the number of transistors which can be packed into an integrated circuit at minimum cost — and therefore the approximate performance (measured in calculations per second) of commercially available computer processors — *doubles* every 1–2 years.

This pattern has been suprisingly closely followed over the last 40 years, as the computer industry has grown from a niche interest to the world-encompassing giant it is today. Think about what it really means: in 40 years — not such a long period, only half a typical human lifetime — we have witnessed multiplication by a factor of around *one billion* of the power of the computer. Most people are nonchalant that while we were talking about computer memory in units of kilobytes just 20–25 years ago, we have moved through the megabyte era to now being able to expect gigabytes in standard desktop or portable computers today (a gigabyte is one million kilobytes). The terabyte era will be with us imminently. Where is this all going?

Kurzweil argues that this progress has been sustained because at each point in time the current generation of technology is used to design the next. Different areas of science and technology interact to mutual benefit: fundamental physics and chemistry, computer science and AI, processor design, genetics, brain scanning and nanotechnology benefit from each other's advances, this exchange enabled by by similarly progressing communications technologies. It can also be seen as the continuation of a much longer exponential trend of progress through human history, where the time between 'paradigm shifts' from one key technology to the next has got smaller and smaller — think of the progress of the cutting edge of communication through speech, writing, printing, telephony, radio and television, email, mobile phones and multimedia internet applications for collaborative text editing or video sharing.

If progress in general, and progress in computer technology in particular, is truly an exponential trend, we must brace ourselves for a particularly significant event coming in the surprisingly near future. The fact is that if some trend is exponential in nature, after a slow start it will reach a point where its ever-growing rate of expansion swamps everything. In this case, the thing which seems due to be swamped is human intelligence itself, as computer-powered AI catches up with and then rapidly leaves behind the power of biological brains. This is the event coined 'The Singularity'.

Despite a makeup which is very different from today's PCs, the human brain can be interpreted as an extremely powerful parallel digital processor, and evaluated in terms of its raw processing capability in operations per second. But with exponential progress, commonly available computers seem set to surpass this level in the near future — the computer on your desk might be as powerful as a human brain in only 10–20 years. The combined power of all computers currently connected to the internet has probably already reached this level. Once computers have matched human processing performance, they will presumably leave us far behind — by factors of billions — not long after.

Leaving aside for later the question of the validity of comparing human brains and computers directly like this, it is quite clear that raw processing performance is only part of the story when it comes to intelligence. A computer which runs as fast as a brain will not be intelligent in the sense we commonly understand it if it just runs word processors at very high speeds. The things that computers can do are already super-human in many areas — no human has been able to compete with a computer in mathematical calculation or data retrieval for several decades. But what is the software that a computer would need to achieve the perhaps human-like tasks — moving through, interacting with and 'understanding' the real day to day world — which have so far evaded computer science?

That is the question which this book primarily aims to address. AI — the research field concerned essentially with concocting 'intelligent' software to run on normal computers — has moved forward rapidly in tandem with the increase in processor performance. While some have spoken of the decline of AI since its heady early days in the 1960s, actually the results of AI research are now commonly in use in all kinds of situations — you only have to look at something like Google, providing anyone with a computer with instant access to the world's knowledge in a way that was imaginable a few years ago. Rodney Brooks has said that AI stops being called AI once it works and the methods are adopted by standard computer science. The AI field tracks the moving target of the hardest remaining problems.

1.1 Robots

Robots are embodied computers — artificially intelligent machines which do not just connect to the world via the abstracted channels of keyboards and screens, but have the *sensors* and *actuators* (motors or other moving parts) which enable them to interact directly with their environments. At the exotic end of robotics are the devices commonly imagined in science fiction: general-purpose mobile machines with wide-reaching intelligence permitting them to exist and operate in an independent, 'human-like' way in the world. Some in AI research work specifically towards such a goal. Right now, in the real world, however, many more mundane devices are also being charged with tasks which require them to understand their environments and act independently. Automated security cameras follow moving intruders; personal computer interfaces interpret their users' gestures or voices; medical imaging devices report abnormalities in scans to doctors; car driver-aids apply the brakes to avert a dangerous skid or park a car automatically.

All of these devices receive information about the world via sensors, be they video cameras, microphones or various simpler encoders, and produce output either as direct action or an interpretation to be passed to a human operator. All of them contain at their core digital computer processors, whose essential task is to *link sensing and action* via computation.

I shall use the term 'intelligent' to describe such robots and devices, though I hesitate because its implications are not immediately clear. Intelligence certainly should not have the meaning which might first come to mind of 'the ability to perform complex computations', since trivial devices such as pocket calculators can achieve that. I have always thought that when speaking about humans, intelligence might be better defined as 'the ability to see truth', to cut through distractions and get to the heart of complex information.

But actually does our understanding of the word intelligence refer intrinsically to our experience of human behaviour, and is it therefore not really applicable to the description of artificial machines? Pinker has described intelligence as 'using knowledge of how things work to attain goals in the face of obstacles', and this seems an appropriate definition. However, he was discussing humans in the context of their evolutionary development into exploiters of the 'cognitive niche' — humans have evolved a generalised and flexible understanding of the physical world to gain an adaptive advantage over the more specialised reasoning capacity of other animals. Could a robot, whatever its particular cognitive talents, be considered to be *intelligent* outside of this evolutionary context? In particular, could we possibly call it intelligent if it continued to serve the interests of its human creators without a care for its own wellbeing?

A more relevant definition of intelligence in the context of robotics is according to Kurzweil along the following lines: 'Using knowledge to achieve tasks in a timely manner'. An important part of this definition is the realisation that speed of operation is important in defining intelligence — a machine which can reach relevant conclusions but take a million years to do so is in fact of no use at all — certainty this type of intelligence would never have evolved in biological organisms as we understand them on earth because it would offer no evolutionary advantage.

Philosophical questions such as the definition of intelligence will at times be relevant to the arguments of this book, but the main focus will be more pragmatic. My aim is to explain the issues involved in enabling real robots and other devices to perceive enough of the world to interact with it and achieve tasks in uncontrolled, everyday scenarios. The term 'interact' is of key importance here, because it implies the aim to construct embodied, stand-alone systems which can operate in real-time in the dynamic, changing world — using computer processors and other materials available today or in the near future. The constraints that these conditions enforce on the types of computational approach that we are able to follow are something that I will look at in depth. In AI we often talk about computation as a resource or currency, available in a fixed amount (depending on the speed of the computer processor available) to be 'spent' as effectively as possible. We will see that some tasks are computationally cheap, and some are very expensive.

1.2 AI for the Real World

Much of the content of this book relates closely to my ten years of both theoretical and practical experience of working with with state-of-the-art research robots, cameras and computers at the University of Oxford, the AIST research institute in Japan and most recently Imperial College London. My research has centred on the problem of autonomous mobile robot navigation, using vision as the primary sensor. The long-term goal of this research field is to produce a vision system which could be attached to a robot dropped, flown, sailed or tunnelled into an environment about which little or nothing is known in advance, and enable it to navigate safely and automatically, building a map as it goes. Today's research robots have achieved subsets of these capabilities in certain simplified environments, but no robot has yet come close to replicating, never mind surpassing, humans' instinctive flair for navigation. The abilities of humans and animals therefore stand as a constant inspiration and in this book I will refer to them often — though let me repeat, in the spirit of the first few paragraphs of this chapter, that the point where robots surpass biological performance even in this domain will probably come much sooner than most people think.

My research lies at the intersection of the young, overlapping research fields of computer vision, robotics and artificial intelligence. These are areas in which many of the most basic issues are, at the start of the 21st Century, still wide open. In particular, progress has been slower than many were led to expect by the rapid early advances made in the early days of computers in the nineteen-sixties and seventies. In this period, robots and AI systems were demonstrated which were able to achieve impressive feats in simplified domains — escaping from mazes, building towers from coloured blocks, playing number games and so on. The methods behind such demonstrations have come to be known as 'classical AI' approaches, in which the state of the robot and the tasks it must tackle can be relatively straightforwardly symbolically represented within a computer and the robot's actions planned using logical deduction as it is commonly understood as the manipulation of statements known to be true or false. The modern extrapolation of such work is AI systems which for instance can now, finally, play chess to such a standard as to challenge and normally beat human champions.

It has become very clear however that this type of approach does not extend in any straightforward manner to enabling robots to complete tasks in natural, *real-world* situations. In the era when a computer program can beat a grandmaster at chess, we are still experiencing great difficulty in building robots able reliably to recognise a person, move across a cluttered room, pick up a piece of fabric, or carry out any number of other tasks which are trivial for a small child.

In order to be able to emulate the actions of a child in picking up a an object seen on the floor, a robot must ask itself and successfully answer questions including: Is that something on the floor? How far away is it? How big is it? What is it made of? *What is it?* Is it safe to touch? Which way up is it? Can I pick up? Is it soft? How heavy is it? There are many questions, each having a great number of possible answers. Classical AI programs, even when running on powerful modern computers, soon become overloaded when faced with such unconstrained complexity.

Chess is essentially a very abstract war game, in which the entire 'state of the world' is specified by the positions of up to thirty-two pieces on a grid of 64 squares. The set of allowed moves is well-defined by the rules; and of course no factors external to the game have any influence. While the number of possible moves at any time is normally in the dozens, many of these can be trivially ruled out as weak and the number of potential candidates for a good move is often low. Even so, predicting the outcome of a certain candidate move a few steps into the future involves a rapid branching of possiblilities which require great computational effort to analyse. Only recently have standard computers reached a level of processing performance where they are able to investigate the outcomes of possible actions in enough depth than they can beat top human players.

There are other games such as the Chinese board game 'Go' (in which black and white stones are used to fight for territory on a 19×19 grid) which have a much larger range of possible moves at any time — and it turns out that computer programs which play Go are still a long way behind the best human players, because they are unable to search through the many candidate moves with enough depth. Humans apparently play this game (and chess) in a very different way, visualising and memorising patterns of territory and playing out long term strategies — a technique that has so far proved difficult to capture in a computer program. By contrast, many games simpler than chess from noughts and crosses right up to draughts are solved problems from the point of view of a computer. The number of possible configurations in these games is small enough that all of them can be reasonably evaluated. The optimal move can be calculated from any position with only a small amount of computation.

On top of the issue of complexity, classical AI methods have trouble coping with the uncertainty involved in dealing with the real world. Before taking any action, a robot must first determine the current state of the world (at least the parts of it relevant to the action), and when this is not provided to it trivially as in board games it must be inferred from data arriving at the sensors it possesses, such as cameras. If the surroundings comprise a simplified block-world of regular planes in bright colours, this inference task might be straightforward: different planes could be reliably segmented in the camera's view based on colour, and their positions relative to the robot deduced from stereo triangulation, for instance.

In real-world scenarios, however, inference is invariably much more difficult, leading to ambiguous interpretations and uncertain state estimates. Human-built indoor scenes contain blank and textured walls, a range of objects occluding each other in arbitrary positions, various light sources, reflections and so on; outdoors a robot might be surrounded by grass, trees, buildings and roads of subtly varying shades. In either case it is likely to be very challenging to segment the scene into different objects, identify what those objects are, and recover their positions relative to the robot. Very often there will be more than one possible interpretation of what is in the scene. Is that vertical feature the leg of a person or of a table? Is that the edge of the road or just a crack in the tarmac? Was that moving object in the edge of the field of view something dangerous or just a shadow? Is that object getting bigger or coming closer? Many of these ambiguities cannot be resolved immediately; it may be that the robot must move to a new viewpoint to discriminate the possibilities, or simply accept that certain things about the world are uncertain at this stage.

Harnessing background knowledge is often the way to way to shed light on ambiguities, though many times background knowledge offers only indirect benefits or has its own inherent uncertainties. Consider the task of reading printed or hand-written text, a problem at which computers have recently started to achieve good results and in which it is now well-understood that background or 'domain' knowledge of language structure plays an important role. We humans find it much easier to read a page of text which is the right way up to within some tolerance than one which is upside-down. Since most of the text we have spent time looking at since we first learned to read has been correctly oriented, our minds have come to expect this and our visual processing when reading has become adapted to searching for patterns of characters with this property.

At a finer level, our ability the recognise the individual characters which make up a word is greatly aided by our knowledge of the language we are reading. Scruffily drawn or even missing letters do not prevent us from reading familiar words because we digest letters in related groups and are able to fill in the gaps. The details of this knowledge certainly appear to be learned since birth, since languages from around the world have completely different letters and writing patterns (some are commonly read vertically). No doubt the adaptations our highly trained brains have made makes our reading more reliable and efficient when we are faced with text which fits the standard set of conditions we have trained with: words and phrases just seem to pop out of the page because we have seen them like that so many times before. However, this does not mean that reading upside-down text is impossible — we can still struggle through this, taking more time and care. We spend more time examining each word to check that our initial hypothesis about what it is is right, and that we have not mis-read a different word. Side-ways text may be a little easier to read; text which is only 45° from upright probably presents less problem still; and text which is only 10° or 20° off upright can be read with almost no problem.

Presumably it would make sense to train a robot with a vision system to read in a similar way — to learn by example that text is upright 95% of the time you want to read it, and that it is worthwhile to develop particularly efficient and robust strategies in this case. When presented with text at another orientation, however, it will have to take somewhat more care, considering a wide range of hypotheses over a longer period before deciding on a final interpretation. The point is that this change of strategy, in this and many other interpretation tasks, is not a binary switch: the ability to interpret and the strategy required change continuously with the orientation and other variables such as the quality of lighting, distance the page is viewed from or text font.

1.3 Strong AI

These concepts of reasoning about uncertain information lie outside the true or false world of classical AI methods. In the years since the weaknesses of classical AI have become apparent, many researchers have gone back to basics and tried new approaches from the ground up in an attempt to provide handles on the issues of uncertainty and complexity which must be tackled to produce intelligent systems which can operate in the real world. In fact, many might have been tempted to give up entirely on the goal of artificial intelligent machines were it not for a growing belief in the scientific community that the operation of a biological brain is in all important respects equivalent to the operation of a digital computer — and therefore that the feats of the only examples of really intelligent behaviour to date, humans and animals, can at least possibly be emulated with computer processing as we understand it today. This is called the 'strong AI' hypothesis.

By describing biological brains and artificial computers as equivalent, I and other supporters of the strong AI hypothesis certainly do not mean that a human brain contains structures anything like the central processing unit (CPU), registers, memory and programs of a modern desktop

PC. The human brain is without doubt a very different unit, comprised of vast bundles of interlinked modules with fiendishly complicated patterns of serial and parallel inter-connections. The assertion of equivalence is only at the lowest level, that the individual neurons forming the circuits of the brain are straightforward computational units, the electrical pulses arriving at and leaving each being related by calculations which can be emulated perfectly by today's electronic digital processors.

This may appear uncontroversial, since when analysed anatomically neurons seem to be nothing but special cells evolved to play the role of electrical connectors, combining the signals at their inputs to produce outputs which are straightforward weighted sums. However the full implication of the strong AI hypothesis is that a large number of these simple calculations is *all* that is going on in the brain — and therefore that all intelligent human behaviour, including the most introspective aspects of the conscious human mind, are consequences of nothing more than the combination of many such trivial calculations. Supporters of strong AI believe that high-level thought must arise from the detail of how the basic processing units of the brain are put together in vast numbers.

Objections to the strong AI viewpoint are well known, including of course spiritual standpoints that there is a 'soul' or something else acting deep within or driving the human mind, as well as scientific theories that the physics of the brain has yet to be understood properties which mean that the computations it can carry out are fundamentally different from those a digital computer can achieve. Such views reflect a natural human reluctance to believe that all the capabilities of the human mind can be explained by a network of trivial processing units. If that is all that is going on in the brain, where does understanding take place? How does consciousness arise? Also, should it not be easy to replicate human intelligence with artificial processing networks?

Dennett, in supporting the strong AI hypothesis, argues that it is easy to underestimate what a large network of simple processors is really capable of doing. When one tries to think of the human brain as a machine, most do not get close to imagining what a machine with the power of the brain must actually be like. It is astoundingly, unimaginably complex; the pattern of connections and responses can encode vast amounts of prior information; there are patterns of signals in different regions which are correlated to and therefore represent the stimuli arriving at the senses, the actions to be performed, and in a more general sense the surrounding world. 'Understanding' occurs in all that detail: particular processing routines are called, particular pieces of internal data changed. Pinker has referred to the human brain as an 'army of idiots'. Each local processing unit or neuron performs a trivial role, but the interaction of vast numbers of them can lead to complex behaviour.

With this viewpoint, creating artificial intelligent devices appears possible but the full difficulty of the task is now very apparent. The roles of the 'idiots' are now played by lines in a computer program, or connections between different processors, each performing a straightforward operation. Turing proved that digital computation is universal: a given calculation can be carried out by any modern programmable computer (with enough storage space). Any computation performed by a network of computers exchanging information can also equivalently be performed by a single sequential processor dividing its time between different, interacting tasks. It would seem therefore that we need not replicate the exact structure of the brain with a massively-parallel digital processing network in order to achieve intelligence — although for practical reasons, this may well be desirable. Given the right program, enough processing power and storage, a computer equivalent in design to the one sitting on your desk should be able to emulate a biological brain and act intelligently.

The perceived failure to date of AI and robotics research to produce many convincing examples of intelligent behaviour I put down to the infancy of the field rather than the fundamental incompatibility of computers and intelligence. Remember that computers have really only been widely and cheaply available since around 1980. Processing power is also a crucial issue. Desktop computers increase in processing power at a staggering rate as the years pass, but they are still apparently behind the human brain when it is evaluated as a raw processing device carrying out a certain number of calculations per unit time using its neuronal hardware (sometimes called 'wetware'). Various researchers have forecast the point at which computers will overtake the brain in this regard to lie in the near future. When it point is reached it will be possible to make a much fairer comparison of the capabilities of computers and the brain.

1.4 Practical Robots

The main subject of this book is not the development of robots which replicate humans or specific animals, either in terms of structure or capabilities. There are research groups around the world working in biologically-inspired robotics, neural computation and related fields, taking brain structures, muscle fibres, retinal patterns and so on as direct guides to the design of robots. Such work, so far, has led to many interesting demonstrations of specific behaviour, but understanding of the higher-level parts of the brain is still at an early stage and replicating through this approach the overall behaviour of any animal, besides perhaps the simplest of insects, has still not been achieved.

Rather, my emphasis will be on approaches motivated by an engineering desire to enable intelligent machines to perform tasks in the real world. While some of the techniques in cutting-edge research robotics are heavily influenced by biology, many more are derived directly from the deep mathematical understanding we now have of the processes of logic and inference. In fact, I think it is becoming increasingly apparent as the level of AI increases that while human and animal brains do many incredible things, artificial computers are capable not only of duplicating their capabilities but also using very different sorts of approaches to achieve sometimes much more — due to their much greater flexibility as well as of course raw processing speed.

While I will mainly discuss the exciting domain of autonomous, freely-moving robots, intelligent machines are not necessarily all robots in the commonly understood sense of the word, since for reasons I have explained it has proved difficult to create fully autonomous mobile machines which are actually useful for something. In the short term, more likely to be practical and useful are various intelligent devices for monitoring, surveillance, human-computer interface, medical imaging and so on.

Ideally the intelligent behaviour of such machines, or the fully independent mobile robots of the future, should be (a) automatic, meaning that they require no human assistance; (b) efficient, so that the processing they need to do to complete tasks can be accomplished without extravagant computing requirements; and (c) robust, meaning that they will function reliably in a wide range of circumstances. While these goals are pragmatic from the point of view of creating useful devices, I have little doubt that pursuing these aims in the development of artificial machines will ultimately shed a lot of light on our understanding of intelligent biological systems. After all, what are humans and animal brains but automatic, efficient and robust processing units with the goal of serving their owners' best interests? It is likely that many of the solutions found by our engineering approach, if they genuinely *work well*, will have much in common with the solutions to similar problems discovered by evolution in the development of biological brains, even if the low-level details of the two domains are very different.

The main research problem from which I will draw examples throughout this book is the question of how robots can be made to move autonomously through an unknown world. The abilities required of an autonomous robot in this area include localisation, map-building, obstacle avoidance and path planning but the whole field can be summarised by as 'spatial awareness'. This is a fundamental problem in robotics and it is the one I have continued to focus on throughout my research career.

Despite the high expectations many people have after seeing the effortless way in which fictional robots in books and films move through their surroundings, the real mobile robots which have to date moved beyond the laboratory to perform useful jobs are severely limited in their navigational capabilities. Some operate in environments such as factories which are controlled enough that they can essentially run on rails, following preprogrammed routes marked out for them. Others need a large amount of human input to guide their actions, such as the rovers Spirit and Opportunity landed on Mars by NASA in 2004 whose movements of a few metres were carefully planned each day by operators on Earth (although in the latter stages of these very successful missions, once the primary goals had all been achieved and exceeded, the robots were given the freedom to explore much more rapidly and autonomously).

Finally there are robots like the low-cost robot vacuum cleaners and lawn-mowers that have recently come onto the market, which move autonomously but in simple patterns and using their sensors at a basic level to bounce off walls and obstacles or stay clear of dangerous areas marked by human-placed warning beacons — these robots do not actually know *where they are* or even if they are revisiting a place they have already been.

1.5 A Domestic Robot

Let us imagine attempting to build a much more flexible general-purpose domestic robot which would inhabit a house and whose tasks might include not just vacuum cleaning but clearing the table after dinner, dusting furniture and tidying away clothes — a robot which could be switched on in the morning and be expected to have carried out many tiresome chores by the end of the day without needing further instruction. Ideally we should not have to modify the house or our behaviour within it to make the robot's job easier — by keeping the corridors clear of clutter, placing barcodes or coloured beacons, or putting objects in places which are easy for it to find for instance.

Domestic robots have been predicted for many years — and there is little doubt that the potential market would be huge. In fact the most common response I hear from people on telling them that I work in robotics is, unsurprisingly, 'when will you make me a robot that can do the cleaning?'. However, as in other application areas, progress toward actual devices which can be bought off-the-shelf has been slow. Of course, in most houses in developed countries today we already have a number of machines which make our lives easier by carrying out tedious tasks with various levels of autonomy: appliances such as washing machines, dishwashers and coffee makers. David Bisset has proposed an interesting distinction which can be made between a domestic appliance and a domestic robot in terms of the space upon which they operate and therefore must to some extent understand and control. In a domestic appliance (and here we mean an automatic device such as a washing machine, rather than one which is human-guided like a vacuum cleaner), the workspace is *internal*, contained completely within the machine itself. A domestic robot, on the other hand, must deal with a workspace which is outside of its physical boundary and therefore affected by many more factors and subject to an uncountably wider range of different configurations.

The cognitive ability required of a washing machine is not zero: modern machines have sensors measuring water temperature, the mass of clothes in the drum, and so on, and control their wash cycles accordingly. However the level of control required, and therefore the amount of thinking to be done, is small because the machine's workspace of clothes packed into a drum is highly uniform and physically constrained. By contrast, to master its external workspace, large in size and filled with unpredictable objects and influences of various types and even people, a domestic robot must have a much higher cognitive level.

In terms of physical construction, an initial design for a general-purpose domestic robot could be arrived at without overt difficulty. Within a single-storey home, a wheeled robot might be sufficient, though far more flexibility would be provided by legs. Prototype human-sized robots which can walk dynamically on two legs over a number of different surfaces, including slopes and stairs, have been demonstrated by several research laboratories, particularly in Japan since the pioneering work by Honda in the 1990s. Earlier walking robots were multi-legged creatures often inspired by insects, and making such designs balance reliably is much easier.

Then the robot would need sensors of various types to obtain raw data about its environment with the goal of finding its way around, identifying objects, monitoring cleaning progress, communicating and becoming aware of unforeseen problems. Sensors which imitate the human senses (including vision, audio, inertial, force, heat and chemical) are available today in many forms and levels of performance and can be straightforwardly interfaced with a computer. It may also be decided to equip the robot with more specialised types of sensor, since there is not necessarily a good reason to restrict it to sensing modalities similar to ours. For instance a commonly used sensor in mobile robotics is a laser range-finder, which emits a harmless and invisible infra-red beam and checks for reflections to determine the distances to surrounding objects. Sonar sensors are also popular in robotics, emitting ultrasonic pulses and receiving information about the environment from the pattern of reflections in the manner of bats or dolphins — in fact sonar has proven particularly important in underwater robotics where vision is only of use over short ranges.

Finally, the robot would need arm-like attachments to manipulate objects and actually carry out cleaning tasks, potentially with various special end-effectors for picking up, sweeping, scrubbing and polishing. High-performance robot arms of various sizes capable of remarkably fast and precise motions have been available for several years, and are in common and familiar use in factories making cars and other goods. The development of robot hands which improve on industrial grippers is a current research area of great interest, the desire being to achieve delicate and flexible multi-fingered motion while keeping the moving mass low. This is being achieved using a number of strategies including indirect cable drives or pneumatically-actuated links reminiscent of biological muscles such as those made by London's Shadow Robot Company.

So putting together the components of our domestic robot could therefore potentially be accomplished today with a high degree of competence, although the bulkiness, weight and power requirements of the robot would no doubt be more then desirable. In fact the inefficiencies of electric motors and the limited storage capacity of batteries are considered by most as major restricting factors in current robot design, and new technologies such as fuel cells or electrically-responsive fibres are keenly awaited.

Building a robot however is really only the very first step. To finish off our robot design we would include some type of powerful computer processor, hopefully of compact size and low power consumption, to serve as a brain, and interface this with the various sensing and action-related components. But what would this processor do? What *software* would run on it to process the information from the sensors and guide the robot's actions? Of course that is what this book is all about.

Let us consider how a domestic robot could move autonomously and purposively around a house. Let us assume that you have just bought one of these robots, brought it home and unpacked it in the living room. You turn it on and the robot opens its camera eyes and stands up. It senses its new surroundings for the first time, and must quickly make enough sense of this environment to begin working in it. It will certainly require a certain amount of instruction from its new human owners before it can go about its duties. To be useful and a real time-saving device however, the robot must have enough intelligence to take most aspects of living and working in a house in its stride, performing automatically and without complaint.

We can compare the situation with what would happen if a new human cleaner had just come to work at your house or office for the first time. Some teaching would certainly be necessary: you would need to explain which jobs need doing both on a regular basis and at certain times, where the cleaning equipment was kept, and probably a number of other specific details — that this door is stuck and difficult to open, or that particular care should be taken in cleaning a fragile object. Beyond this, however, you would not expect to have to explain much else. You would certainly not need to give the cleaner a map of the building. In a large house you might need to give brief verbal directions about the positions of various rooms — the bathroom is at the top of the stairs on the right, etc. But subsequently you would not need to remind the cleaner about how to get around.

This ability of a human to understand effortlessly the layout of a house is thanks to a combination of rapid learning and a huge store of prior knowledge and experience. A human knows that in most modern western houses there are one or more bathrooms, that the hall is the space next to the front door, that in a two-storey house the bedrooms are usually upstairs, and that the kitchen is the room with a sink and washing machine. But at a lower, instinctive level a human also knows a vast number of 'common sense' facts about buildings, life within them and the way the world works in general: that doors normally have handles which are turned to open them, that a wet floor may be slippery and requires more care to traverse, or that objects placed near to the edge of a table are in danger of falling off. In the human mind, most of these facts will have been learnt through the long trial-and-error experience of childhood, teaching from elders and culture, and some of the most important are even more deeply ingrained as inherited instincts.

If our domestic robot, standing in a new house for the first time, is to function autonomously it must certainly have a similar library of background knowledge to draw on — whether this knowledge is pre-programmed during the robot's construction, or learnt during its previous experiences. This knowledge must then combine dynamically with the new information arriving at its senses to enable it to make sense of the world around it and decide on its actions.

A large amount of the information collected by the robot's sensors and manipulated by its processor will be immediately discardable because it is not relevant to the robot's existence and tasks. High level irrelevant facts such as that a bird flew past the window, or that there is a certain headline on the front of today's newspaper, could certainly be discarded once identified as such (we will leave the question of how these events are recognised). But at a much lower level there is a vast amount of detail in the uninteresting patterns of raw data captured by visual, audio and other sensors (the rippling of sunlight on the floor coming through the tree by the window or the hum of the washing machine) which should be discarded before troubling the flow of the robot's main processing and activity. Other data will need to be processed for immediate reference, but afterwards can be safely forgotten: a person entering the room within which the robot is working will need to be monitored to avoid a collision, but this event can probably be erased from memory once it has finished.

On the other hand, some of the newly sensed information will be particularly important, and itself need to be stored in long-term memory for future reference. Certain consistent patterns of new data should even permit the robot to learn new generic rules about how the world around it works. For instance, after spending a few weeks in the house the robot may detect that every Sunday there is a mess near the back door thanks to the children returning from their football practice with muddy boots. The robot should be able to add a particularly thorough scrub of this area to its regular cleaning schedule.

1.6 Maps and Navigation

Now in order to find its way around the house, and be able to revisit places it has been to before, the robot needs to lay down a particularly important reference in its long-term memory, telling it how the various rooms are laid out and how they fit together: it must make a *map* of its surroundings. The map it should perhaps ultimately aim to construct would at the essential level look like the plans used when the house was first designed and built. These were 'blue-print' line drawings showing projections from three directions at right angles of the geometry of the house's outer and inner walls and permanent fittings to give familiar top, front and side views. If these plans were available to give the robot when it first entered the house, its work would no doubt be made easier.

However, plans are often lost, house extensions and modifications are made, and walls lean with age. More significantly, a house does not contain just the bare walls and permanent fittings of its blue-prints: it is full of furniture, curtains, shelves, and a host of objects which are moved around from day to day. The robot must therefore inevitably have the ability to make maps for itself, not just of the house's basic structure but of the locations of all the things which are relevant to its tasks and the safety of itself and the other occupants. Creating and maintaining an accurate map would give the robot geometric mastery over the house and is certainly a desirable state of affairs. It could instantly calculate the exact distance and shortest path between any two points it decided to move between, or work out exactly how much cleaning fluid was required to do all the floor surfaces.

For a human to make an accurate geometrical plan of an unfamiliar house would require a lot of effort, and some specialised tools for measuring distances and angles. However, of course humans do not need to go to such lengths to learn enough about a new house to be able to move sensibly around it. We find that we are able to capture the essential layout of a simple building without really trying, and that the quality of this knowledge improves with time spent there. Most people could sketch a plan of their home or other very familiar building from memory, showing the layout of rooms and how they fit together, the locations of doors, windows and stairs — though probably not annotate this plan with precise measurements. In the case of a less well-known building, such the the home of a friend, an attempt to draw a plan may result in a drawing which is vague or perhaps wildly inaccurate.

So a human has an intuitive grasp of the layout of the rooms around a house, but not necessarily an accurate concept their geometry. But the truth is that a human does not need such detail to be able to move safely about. Instead of concentrating on precise measurements of distances and angles, humans are adept at absorbing the gross structure of a room, picking out landmarks, interesting objects and features, and seemingly relating their positions in an approximate way. We know that a certain room has green walls, a large window on one side with a view of the road, a television next to the sofa and a door in the opposite corner. We certainly seem to have a good rough idea of the size and proportions of the room, but are not very sensitive to the precise values of these. Rather, a quick glance around the room to identify major landmarks is what is required to confirm the identity of a location.

Humans are also able to navigate through very large environments for which they would be even less likely to be able to draw representative plans with any large degree of geometrical correctness. Sitting at home, you would probably not be able to point accurately in the direction of a place you visit often on the other side of town, such as a particular shop (though this may be easier if your town has a grid-based layout), or estimate the distance to it with greater than perhaps 50% accuracy. However, it would be straightforward to leave home and walk or drive reliably to the shop without needing to look at a map or use a compass. Humans seem to achieve such feats with a style of context-based navigation. Rather than keeping in mind the geometry of the whole town or city, we know how to get from the current location to the next on a familiar route, from there to a subsequent waypoint, and so on, guided by the recognition of locations and specific landmarks. Of course, navigation around a town is greatly facilitated by the fact that one's freedom of movement is generally restricted to the network of roads and paths. Navigation choices only need to be made at intersections, so there are fewer ways to go wrong. Straying from the network, for instance by striking out across a large park as a shortcut, leads to more difficult navigation thanks to the increased freedom of movement.

These types of navigation used by a human appear to be in many ways *topological* rather than *metric*. Topology is the branch of mathematics which studies the properties of geometric figures or solids that are not changed by transformations such as stretching and bending. In topology, a narrow rod is equivalent to a sphere or cube, and a ring is equivalent to a tube, since each of these shapes can be changed into the other by a continuous deformation process. The properties of a shape important in topology are the numbers of holes and inter-connections between different parts, since these are not affected by smooth transformations — and not the precise measurements and angles describing its dimensions which we refer to as metric quantities.

A topological *map* therefore stores only the information about which places are connected to others, while neglecting metric details of distances and angles. Such maps are also sometimes known as 'qualitative', as opposed to 'quantitative' metric maps. A familiar example is the London Underground rail network map — considered by many as a triumph of clear graphic design — or its counterparts in other major cities. The role of these maps is to display the inter-connections between stations in a standard, simplified way: stations are displayed as being equidistant along coloured rail lines, and the lines follow paths made up of smooth straight-line segments which do not closely represent the twisting and complicated shape of the real-world tracks at all beyond the points where they intersect.

A topological map of connections between locations becomes useful when the mapped locations and perhaps also their inter-connections can be labelled with names which describe them and allow them to be recognised: the map acquires *semantic* value. It can then be used to guide navigation by allowing routes from one place to another to be inferred: it tells us that in London to get from Victoria Station to Tottenham Court Road one must go two stops on the Victoria Line to Oxford Circus, change trains, then go one further stop on the Central Line. When actually using this information in practice, recognising a particular Underground station from the train is usually trivial thanks to signs and audio announcements, and so following a route planned on the map is straightforward.

A less refined example of a semantic map is a sketch representing the route to a friend's house drawn upon hearing brief telephone directions. This map might be drawn to a rough scale representing approximate distances of 'around a kilometre' or 'five minutes' walk', but would also likely have breaks or strange contractions of distance when the edge of the page was getting too close and so on. Much more important to the map's usefulness would be semantic labels, indicating the location of landmarks such as particular buildings but also the topological structure of the road network. We understand what is meant by 'the statements 'turn left at the T-junction' or 'go straight on at the crossroads', can draw parts of our topological map representing these structures and most importantly can recognise when we have reached these locations in the real world. On the other hand of course, we have all had experiences of getting lost when trying to follow oral directions and maps of this type, and this reflects the fundamental difficulty of transmitting the map that one person has in their head to another person via oral descriptions and sketched symbolic representations. This translates directly into the difficulty researchers have experienced in programming robots to perform human-style navigation.

Topological navigation relies on the ability to recognise locations from their semantic description in the map in the presence of a lot of metric uncertainty. The map says to 'go straight ahead for around a kilometre and turn left at the supermarket' — and so to follow these directions one must check every building passed between perhaps 0.5km and 1.5km to see if it is a supermarket. The map only works if the reliability of the map user in recognising the supermarket is very high. Problems will occur either if another building is falsely identified as the supermarket, or if the actual supermarket is missed. However, humans know that supermarkets are buildings which are easy to spot, since they will be clearly marked in a familiar way and much bigger and rarer than standard houses.

Semantic, topological navigation in humans works in part because humans have phenomenal powers of recognition, of objects and locations. A human can trivially recognise a door or a window, item of furniture or other familiar object from most normal angles and ranges of view (though not from very far away, when objects appear too small for detailed visual information to be extracted, or extremely close-up when only a fraction of the object may be visible and may appear distorted by the proximity). We can recognise classes of such objects as well as specific instances: a chair is a chair even if we have not encountered one of this particular type, size, shape and colour before. Human visual recognition also copes astoundingly well with differing lighting conditions; the amount of ambient light falling on an object can vary by many orders of magnitude and we almost do not realise it because it is still trivial to recognise it. It is even often possible to recognise an object which is only partly visible, thanks to occlusion by some other closer object.

But object recognition is a currently a difficult task for artificial computer vision systems, and the recognition abilities of even the most advanced research robots are still quite limited (and other types of additional sensors the robot might carry do not help much). Humans can open their eyes and be presented with almost any type of object they have seen before and immediately say what it is. A robot, even if it has been pre-programmed with a vast library of information about objects, faces a huge challenge in accomplishing something similar. Based on the uncertain data available in one or a few images, combined with similarly uncertain context-based prior information, and reasoning about all the possible positions, orientations, occlusions, lighting conditions and variations in type of the objects it knows about (with all the inherent ambiguities), it must make a *rapid* decision about the identities of the objects present — rapid if this decision is to be of any use in guiding the robot's immediate future actions. In the next chapter I will look in some detail at how computer vision object recognition systems of today actually work, and how they might be made to work better in the future.

In the meantime, this means that as a robot moves around a house, it cannot enter a room, have a quick look around as a human does (acquiring images in several directions) and identify immediately a 'feel' for the room via a set of landmarks objects which characterise that particular place. This makes the type of topological navigation used by humans impossible except in certain simplified scenarios. Building and using topological maps has been demonstrated by mobile robots moving around corridor networks made up of clean, straight lines, right angles and flat floors. The robots were equipped with range sensors which allowed them to extract the local shape of the nearby area as they moved and classify each location as 'straight corridor', 'T-junction' or 'crossroads'. These locations, assuming that all were correctly identified, could then be joined into a topological map. These demonstrations worked only because the level of variety in the environment was very low, and its uniform nature meant that it was relatively straightforward for the robot to identify the local scene structure.

We can see that a map of a more general and complex scene will become increasingly useful to a robot if its semantic information is accompanied by high quality metric data. Suppose that the 'around a kilometre' distance to the supermarket on the route to a friend's house is actually measured more accurately and the map now indicates that this distance is 940m, and that a robot is now charged with following the map. The robot will almost certainly have the ability to measure its local motion at any time with a reasonable degree of accuracy using internal 'odometry' which counts the number of times its wheels turn and multiplies this by their circumference, like the odometer in a car which counts up the kilometres it has travelled. A legged robot would instead keep track of the number and length of steps its legs make. Humans too have some level of odometry, since it is possible to traverse short planned distances with closed eyes and keep and idea of where you are, although the level of accuracy here is far below that a robot can achieve. Now, within some margin determined by how much the robot trusts the distance information in the map and its assessment of the reliability of its own odometry measurement, it can narros its search for the supermarket.

For example, it may assess the map to be accurate to a tolerance of plus or minus 10m over this distance (perhaps thanks to an accuracy specification provided by the maker of the map, or the robot's previous experience with maps of the same type). Similarly, it may know that its odometry has an uncertainty of plus or minus 30m during such a journey — uncertainty in local motion sensing such as odometry is inevitable, since simply counting the number of wheel turns or steps taken as a way to measure distance does not take account of unpredictable factors such as slippage. Combining these two sources of uncertainty gives it a range of plus or minus $\sqrt{30^2 + 10^2} \approx 32$m within which the figure reported by the odometry should agree with distance moved across the map — note that I will look in Chapter **??** at why the uncertainties should be combined in this squared/added/rooted way.

Therefore, starting its journey along the road, it can ignore all buildings passed while its odometry reads less than 908m, then start searching for the supermarket as it continues to move until 972m. If it does not find it before 972m, it will know that it has missed it and can potentially turn around to go back and search in more detail. The chance of making a mistake is now much reduced because it is unlikely that there will be two large, supermarket-like buildings within an 64m stretch whereas this was a serious risk within the previous 1000m search stretch.

So a good quality metric map may include all the semantic information contained in a topological map, but also offer extra value in the form of measurements. We should also remember that the goal of building robots is not necessarily to duplicate the abilities of humans. Since computers are very good at storing and manipulating numbers with extreme precision, surely a robot can and should do better than topological navigation? When all is said and done, a metric map is the 'correct', accurate way to represent a location, and it offers possibilities not available with only topological representation. A robot (or human) with a metric map and some way to measure its local motion (odometry, perhaps in combination with a compass to measure direction — in the human case odometry could take the form of timing combined with a knowledge of normal walking speed) can strike out across a wide region like the park in the middle of a city within which there are few or no landmarks, but still have some way to calculate the right direction to head in and predict the distance to a distant waypoint. Robots are actually likely to be able to do this much better than humans because they can trivially handle the explicit calculations which are required.

1.7 Intelligence as a Modeller and Predictor

We have started to explore some of the aspects of the 'thought processes', implemented as computer programs, which a robot needs to exist autonomously in the world. We have begun to see in particular that in a sensible and efficient and approach to moving reliably and safely through any real, complex environment, a robot's progress through the world is accompanied by the propagation and evolution of internal estimates about its state and that of the world around it. A robot which moves around and interacts with the world must carry out its processing in real-time, since the results of calculations must be available immediately for use in decisions about what to do next. This real-time requirement is what forces a sequential approach to the propagation of estimates through time. This is true not just in localisation and mapping, but in any real-time intelligent system.

Crucially, we have seen that the robot had to maintain and perform calculations with not just point estimates of the aspects of the world in which it is interested, but a representation of the uncertainty in these quantities. This uncertainty is not some inherent property of the the world, but simply reflects the lack of perfect knowledge the robot inevitably has when it has only had limited experience of the complex world and the way in which it interacts with it. We would expect that given enough observations of a particular domain, the uncertainty in the robot's world view, at least with respect to the static components of the environment, would tend towards zero. This has been both theoretically and experimentally shown in robotic localisation and mapping: a robot which spends many hours moving around a small environment, making very many repeated measurements of the locations of features from different posititons, can eventually become arbitrarily certain of their locations. I should qualify this slightly by remembering that the concept of a feature itself is just a model or approximation to any real scene geometry and therefore the true limit in the uncertainty of the location of a feature which the robot will tend to over time will be determined by how well the feature representation fits reality — though with sensible feature choices this fit can be very close.

One way to describe a robot's computational process is to think of it, as Steve Grand has suggested, as an 'inner narrative' which the robot develops and continuously tells itself — its own internal version of the structure and happenings of the world, including how the robot itself fits into it. The narrative is the robot's 'personal' view, starting from a point reflecting its experience plus whatever prior information it might have and then evolving in reaction to its actions and the things it senses. An important part of this idea of narrative is certainly a growing set of memorised information about what is where and the properties of objects, materials and sensors. However, it also has a temporal component representing its understanding of the evolution of processes through time. If the robot were to turn off some or all of its sensors the narrative would still progress through time, but in a way which would gradually or more quickly diverge from the truth — in SLAM, this is what happends if a robot proceeds using only odometry, its positional uncertainty growing steadily with distance moved. The robot must use relevant sensory measurements to update or *tune* its narrative, by comparing what they report with what was expected by its internal representation.

This concept of inner narrative brings us back once more to a point of comparison with the operation of the human mind, where something similar certainly seems to be going on. We interact with our environments through a limited set of senses and actions, and can never know all the details of even the smallest real environments. Instead, we tell ourselves a story. The brain builds up and drives a kind of simulation of the world, accurate in some respects and looser in others, which is tuned via real sensory interactions. Dawkins has recently written about the role of the brain as a virtual reality simulator of the aspects of the world relevant to one's existence. To decide what to do next, we visualise or *imagine* the outcomes of potential actions, based on our current state of knowledge — effectively simulating them internally.

These visualisations might relate to very short term planning, such as to decide how exactly to place the next footstep when walking on rough terrain, simulating the answers to questions like 'Will this rock move?', or 'Is it likely to be slippery?'. Such predictions might not rise to the level of what we would normally consider conscious thought, but doubtless must occur somewhere in the brain. At the other extreme, one might imagine the consequences of a decision about whether to accept a job offer across many years via much explicit introspection. In either case, making the decision requires a prediction of the potential outcomes of the action, and this prediction is necessarily grounded in a model held in the brain of the world and how it works.

Dennett, in also advocating this view, has said that the main job of the brain is to 'produce future' — it is an organ of modelling, simulation and prediction. The predictions it is able to make are not limited to the passive dynamics of a rock or thrown ball; they also cover active, *intentional* agents, most importantly other humans. Predicting the actions of another human requires an understanding of their desires and goals, which might seem unfeasibly difficult — but it is made possible by empathy. We assume that in general other humans have the same basic drives as ourselves, and can put ourselves temporarily in their shoes to guess what they will do next. You might be able to predict the position of a friend to phenomenal precision a whole week in the future for instance because you understand his normal routine or the things he likes to do. This process can still be considered however as predicting the future by winding the clock forward on an internal simulation, where this simulation now includes the other agents and estimates of their behavioural traits.

An inner narrative, whether in a robot or a human, is nothing but an evolving personal model of the world. A supporter of the strong AI hypothesis such as myself, who believes that the operation of the brain can be viewed as equivalent to that of a network of artificial digital processors, must see this model as a set of information which can be represented by numbers and mathematical formulae. The new aspect of these concepts of inner narrative and internal simulation which I want to bring to the fore in the context of real-time robotics in this book is the role of uncertainty. As we have already started to see, recent work in robotics and computer vision has come to consider working with uncertainty as the fundamental basis for all kinds of real-world inference. A robot can use data to infer an estimate of some property of the world, but this estimate will likely be of little use without an assessment of its uncertainty. This uncertainty measure is the basis for the essential processes of combining different types of data or setting thresholds for decisions.

Specifically, the most successful approaches to real-world robotics from recent years are grounded in Bayesian probability theory. Probability theory is the same branch of mathematics we are all familiar with to some extent from gambling odds or weather reports. In recent years however its generality, when interpreted appropriately, as *the* mechanism for dealing with all types of uncertain information has come to be increasingly widely accepted. The interpretation which permits this very general use of the theory is known as the Bayesian viewpoint: the idea that probabilities *do not exist* as real quantities in the world, out there to be discovered like frequencies, but represent only the subjective degree of belief an observer holds in a particular hypothesis based on the information with which he has been presented. Two different observers may indeed completely sensibly hold different probabilities for the status of some aspect of the world, thanks to different experience or prior information. A Bayesian observer uses probability theory as *extended logic*: a tool for making inferences from data, except now the degree of belief in a proposition can vary smoothly between zero and one and not be constrained to the strict true or false values of the Aristotelian logic one is familiar with from logic puzzles.

This ties in perfectly all of the discussion we have had about what a robot must achieve, which is to build a model of aspects of the world around it based on the information it gleans, and represent and propagate the uncertainty in this model. Bayesian probability theory provides the sound framework necessary to combine information from arbitrary sources into a view of the world in which uncertainty is explicitly represented. The effect of every small new piece of information gained from a sensor is to reduce this uncertainty in some way, and probability theory, if used properly — and there are still issues in understanding how this should be done in many scenarios — can tell the robot exactly how to do this.

1.8 Real-Time Processing

However, there is a catch, which you may have already forseen. The robots and other systems in which I am interested in this book are embodied, real-time systems interacting with the dynamic world and with limited processing resources. As data streams into a robot's sensors, it would be theoretically possible via probability theory to compose that information into a consistent world picture, cross-referencing every small fact to gradually converge towards a comprehensive representation. But composing all of this information requires a great deal of computation — and at some level of scale and precision of representation, more than any real robot can possibly be endowed with. Instead, the robot must make choices about which parts of the incoming data and its stored world information are relevant to its current situation and future goals, and neglect the rest. This issue is known in artificial intelligence research as the *frame problem*.

Daniel Dennett illuminated the frame problem in the context of a hypothetical robot, required to act quickly in order to replenish its diminishing power supply but surrounded by potential hazards. A rushed decision could lead to its downfall at the hands of some unforeseen danger — it might charge across the room towards its power supply without performing enough visual processing to check sufficiently for obstacles and crash into a wall. On the other hand, over-contemplation could leave it paralysed by indecision — its power supply would trickle down to zero as one by one it computed and predicted the outcomes of possible eventualities. Inevitably, after some consideration it must just accept some level of risk and go for it. Some data, eventualities and details must simply be *ignored* rather than thought through and accepted or rejected, But which? Real embodied intelligence requires the ability to dismiss red herrings rapidly and focus on the important problems. Dennett concluded that 'a creature that can solve any problem given enough time — say a million years — is not in fact intelligent at all'.

The frame problem is all about the assumed context to a decision. Asked the question 'where do you come from?', you might decide to respond with the name of your country if meeting someone when abroad, but with the name of a town when talking to someone to whom your nationality was assumed to be clear, or even with the name of a specific if you thought the person would be familiar with the town. The 'frame' of a decision is defined by the decider's character, experience and goals.

An example of the type of decision a navigating robot would face involves the amount of detail to put into a map it is building of a room. Should it extract just a few point landmarks, or aim to build a much denser representation with textured planar facets closely modelling the shapes of objects? The answer to this question depends on many factors: the goals of the robot (Does it just need to navigate swiftly through this area or does it need to search the room in detail for a particular object? Is it likely to come back this way again later?); the prior knowledge it has of scenes of this type (Can it reasonably assume that there are no glass doors in this building and therefore that all obstacles will show up in its cameras?); the sensors it has available; and importantly the computational resources it has available to apply to the limited time it can spend in this area.

The frame problem is not an issue that can be *solved* by some clever technique. It is a real-world restriction on computation which seems to require a robot to be even more intelligent to be able to solve everyday tasks than one might have initially thought. Even if it had a perfect data-handling and modelling method, able to file all information from its sensors away in the appropriate place for future reference, it would not be able to use it because the computational cost of referring to this ever-growing database always becomes prohibitive at some point. Instead it must somehow know in advance, without analysing all of the data available, which parts are likely to be the most important in any particular situation. It must also selectively forget many past experiences while holding on to key information.

How do humans cope with this problem? It seems to be via a huge library of experience, prior knowledge and instinctive habits which give us an intuitive ability to label blocks of data and focus our attention appropriately. We just 'know' that when the sun goes behind a cloud this is not a significant event which requires us to alter our plans of action, despite the major change to the visual field it causes. We would often not even ‘notice’ it at the conscious level of thought, although lowel levels of brain activity will certainly have reacted and adjusted to it. The human vision system especially has been referred to as a ‘bag of tricks’, specialised tools for common jobs relating tightly or loosely to each other and combining in as yet poorly understood ways.

1.9 Under the Hood of Research Robotics

So does this mean that we are doomed in trying to use sound theoretical and mathematical principles to build useful real-world robots? Is the only hope to try to emulate directly the way the human brain works with neurally or genetically-inspired methods which learn about the world in ‘natural’ ways but whose inner workings are no more easy to interpret (or *control*) than the operation of a real brain? I still feel that this is not the case. In fact, I predict that the power of the methods recently coming into favour in AI research will extend to give a solid theoretical basis for most aspects of the development of real-time robots, and in turn that this research will have great consequences for the study of the human brain.

The main concepts I want to examine in the rest of this book are *modelling* as fully-acknowledged approximation, and how this fits smoothly into the use of Bayesian probabilistic inference to manipulate uncertain information. A model is an abstraction of a complex real-world entity into a mathematical form which is much more straightforward to represent and use. We will see that almost all of the concepts we are familiar with from daily life can be viewed as models. I will argue though that whenever an abstraction is made, it is important to characterise the degree of approximation and therefore uncertainty in the model, since this is critical to its value in making decisions, and that the correct way to do this is using probability theory. In fact there is recent evidence, for example from the striking research of Daniel Wolpert and colleagues, that humans use probabilistic mental processes during sensorimotor control of the eyes, head and arms during day-to-day tasks, and it would be natural to expect more examples of this to be discovered soon.

I will look at how models, as abstractions, become increasingly powerful when combined in hierarchies. Low-level abstractions (like points and lines) build into mid-level (e.g. surfaces) and high-level (e.g. buildings) semantic representations which can be manipulated efficiently without needing to know about all the detail below. As models become more abstract, it becomes easier to see how a computer, or a part of the brain, can reasonably manipulate them usefully with the straightforward calculations possible by lines of a computer program or small blocks of neurons. As Dennett has argued, *understanding*, if it makes sense to use this word in the context of a robot, comes in the detailed operation of software representing these models at different levels of abstraction and managing it in a form which can actually be used to do something useful.

We will look in some detail 'under the hood' of a robot at how such hierarchies can be realised in a computer via tools such as object-oriented programming, and try to give the reader a real feel for how the discussions of methodologies relate directly to what can be achieved by programming a desk-top computer. Much work in robotics and AI is concerned with developing *algorithms* — mathematical recipes, implemented as computer programs, for achieving particular tasks in sensing, inference or action. A typical robotics research paper will present a new algorithm for tackling a problem, explaining the theoretical basis for the method, describing it is potentially better than previously-published alternatives and presenting results which ideally directly show the improved performance of the new approach — where the improvement could be in terms of better capabilities or the same capabilities with reduced computational cost. Discussions between researchers are normally at the level of algorithms rather than the detail of computer code — a good algorithm will prove useful for many different robots in varying scenarios.

I will examine how a probabilistic model relates to a simulation in the sense that we commonly understand the word, and argue that probabilistic simulations will become the widespread engines behind all types of robotics and intelligent inference.

Finally, in what I consider the most important part of the book I will look at the specific issues facing real-time systems. Active methods, which I have already mentioned with reference to visual SLAM, are purposive, or task-driven strategies which refer to the goals of a robot and the current status of its world view in guiding sensors and processing data — they involve choices about what to do when, which sensors to point where and which data to concentrate on, often guided by the predicted outcome of actions. These methods have the potential to be much more efficient than the contrasting passive approaches which treat all data and time-steps equally — and are surely the basis for coping with the frame problem.

Modelling and probability theory already start to give a handle on the frame problem because they can provide an understanding of how abstraction of the world at increasingly high levels into representations which can be efficiently manipulated relates to the uncertainty in any conclusions which can be drawn. Active techniques can be put on a firm theoretical basis within the Bayesian framework by extending it with decision theory and Shannon's information theory, natural additions to probabilistic inference. Information theory is the study of the information content of data, and allows real metrics to be placed on the value of measurements in reducing uncertainty. Decision theory provides a basis for making intelligent choices based on evaluations of the penalties or benefits associated with uncertain predicted outcomes, and can really answer questions about balancing risk and deciding when to 'go for it'. Both of these theories have as yet been seriously under-used in real-time robotics.

In particular, I will look at how these methods will lead to an understanding of whether there is a point in certain situations where it no longer becomes useful at all to use probabilistic theory. Given an image of a scene, for instance, the active approach a robot would take is first to decide what information needs to be extracted, then search the image selectively for the answers to specific questions, and this methodology can be characterised as 'top-down' processing. The contrasting passive or 'bottom-up' way to proceed would be to perform some kind of processing uniformly across the image before considering what information was required from it.

Some types of bottom-up processing are very computationally 'cheap', meaning that they can be achieved with few CPU operations. For instance detecting edges and corner features across a whole image, a very useful pre-processing step for many visual tasks, can today be accomplished by a standard processor in fractions of a millisecond. Making the probabilistic calculations to decide in which parts of the image to try to detect features to reduce uncertainty optimally might be more computationally expensive than just detecting features automatically across the whole image.

In the human visual system there is certainly a lot of pure bottom-up processing ('early vision') which is carried out continuously in the primary visual cortex as a pre-cursor to all subsequent processing. For instance, the brain has structures which have been convincingly shown to perform the roles of oriented filters at different scales and locations across the retinal image which are directly analogous to edge detection in computer vision. But at the other end of the scale there are certainly behaviours at higher levels which are top-down in nature — the frequent purposive saccades of the eyes give a direct insight into this much more selective style of vision, in which short and long-term goals guide the focus of visual processing.

In both the human visual system and real-time robot intelligence, therefore, it seems that there must be a point where bottom-up and top-down processing meet. True understanding of how to cope with the frame problem will come from a theoretical description of this meeting point, which must arise at a threshold determined by the crossing of the computational costs of the contrasting approaches. I certainly do not have all the answers to how this might be found (it will surely lie in a different place in every different task faced by a robot), but I will discuss the general principles of how I believe this field of research will develop.

Before continuing to look in Chapter **??** at the concept of modelling and how understanding it informs progress in real-time robotics and computer vision, in the next two chapters we will take a look in more detail at how real AI systems work, first by returning to robot localisation and mapping, and then by examining the problem of object recognition using computer vision.