Robotic Vision

Lecture 1

Introduction

Welcome to this first lecture in the Robotic Vision course, in this first lecture we are going to recap about robots, how they work, and their need for senses. And then we're going to talk about senses more broadly the sort of senses that animals and people have and then we are going to focus on sort of senses that robots have perhaps things like GPS but in particular we are going to focus on why it is that vision makes sense as a sensor for robots, so my job in this lecture is to make the case for why vision is a great sensor for robots.

Robotic Vision

Lecture 1, Section 1

Robots revisited

Before we start talking about robotic vision, it's going to be very useful just to revisit some fundamentals of robots, and in particular, definitions of what a robot is.

For those of you who participated in the first online course, we presented this definition of what a robot is: It's a goal-oriented machine that can sense, plan and act. And the important keywords are highlighted here in red.

A more casual definition that we introduced for a robot is that it's a machine that can move—either itself or perhaps its hand—from place A to place B. So we might move its hand to pick up an object at place A, and put it down at place B; or it might movie its entire body. It might move along a corridor from place A in the corridor to place B in the corridor.

Let's now look at some of the keywords in the first, more detailed definition, and one of the key concepts is sensing.

A robot fundamentally is a machine that can sense its environment. The sorts of things that it might want to sense is: where is the object that it needs to manipulate—where is the thing that it needs to pick up. Another thing that it might want to sense is where is the robot itself. A mobile robot in a corridor situation; whereabouts in the corridor is the robot? This is the classical 'where am I?' problem of mobile robotics.

Another thing a robot needs to do is to plan. So from its sensors it knows where it is; it knows where the things that it needs to work with are; it has got a goal, a thing that it wants to achieve. So it makes a plan of where to move from its current state to its goal state.

And the final part of the process is to carry out some action. Given that I've got a plan, now I have to carry out the step of the plan. I grab the object or I move myself from place A to place B.

Queensland University of Technology

Robotics Vision

Lecture 1, Section 2

Senses

When I was a young person at school they taught us that we had five senses, but I would argue today that they have overlooked a very important sixth sense, which is the sense of balance. The sense of balance is absolutely essential to upright walking animals like human beings, and we achieve balance using some senses that are located in our inner ear, in the vestibular apparatus where we have the equivalent of accelerometers and gyroscopes that provide the information to keep us upright. Balance is absolutely critical and the human being's sixth sense.

Other animals have evolved different sensing modalities to those the humans have. For instance, small flying animals like bats have the sense of echo location; they send out pulses of acoustic energy and listen for the reflections and they use this when they are flying at quite high speed at night to find obstacles that they would want to avoid and to find insects that they would wish to eat.

Sharks create an electric field around their body and then they can sense disturbances in their electric field, which are caused by fish swimming in close proximity to the shark's body, and this is really useful for them in order to detect sources of food.

Finally, pigeons have got an innate sense of magnetism. They have magnetic senses in their heads that allow them to tell where north is and this is absolutely essential for them in their flying long distances, an essential navigational aid.

Now for us, it's almost impossible to imagine what a sense of 'northness' feels like, but pigeons have that sense, so there are many, many, different senses that animals have evolved over time to help them perform the essential functions of life.

Robotics Vision

Lecture 1, Section 3

Why not use GPS?

When we start talking about robots and the need for sensors, the question that many people ask is: 'Why don't we just use GPS, surely that's enough?'. GPS is everywhere—it's in our phones, it's in our cars … it tells us where we are. So let's talk a little bit about GPS.

Imagine we have a robot on the ground and it's got a GPS receiver in its head. In the sky are a number of satellites in low earth orbit, and they are all sending radio waves which the GPS receiver in the robot's head can pick up.

Using the time that it takes the radio wave to travel from each satellite to the robot we can work out the distance from each satellite to the robot.

If we know enough of these distances—we know where the satellites are in space—then we can work out where the robot is. That's the fundamental principle of GPS.

Now these radio waves are at a very high frequency; they are around about 1.5 gigahertz, and that's significant and we'll talk about why that's significant shortly.

Imagine now we're looking upwards from the robot. We're looking up to the sky, and there are a number of satellites that we can see. We only need four satellites in fact, to work out where we are on the planet. That's the minimum requirement in order to get what's called a 'fix'.

Now imagine that our robot is in an urban environment. Imagine we're on a street in Manhattan or something like that. So what happens now is that some part of the sky is obscured. We can't see all of the satellites. Now we can only see two satellites and that's not enough to obtain a fix. This is a very common phenomenon, that's why we have problems with GPS in urban environments. But it occurs in other environments as well. It can occur perhaps in a very, very deep mining pit where the walls of the pit obscure a significant fraction of the sky.

Now imagine that we have our robot and it's at some big industrial complex. There are large walls and chimneys and structures made out of metal. Now what happens in this case is that the signals from the satellite may not travel in a direct line to the robot. They may bounce off some of these metallic structures before they hit the robot. The problem with this is that the signals have travelled a longer distance than the actual distance between the robot and the satellite. The path length has increased through what's called 'multi-pathed reflection'. So the GPS receiver within the robot doesn't know this, so it will come up with an erroneous estimate of where the robot is located.

This is, again, quite a common problem with GPS next to a big structure that reflects radio waves. The GPS estimated robot position can be significantly in error.

Now let's consider another scenario. Consider that our robot is underground, perhaps doing some mining work, a really important domain for robots. Now the problem that we have here is that the radio waves from the satellite cannot penetrate the earth.

Another application where we might want to use robots is underwater. And again we have the same problem as we have in the underground case. The radio waves at 1.5 gigahertz cannot penetrate very far into water. They perhaps penetrate a few millimetres. So a robot that's at any depth can make absolutely no use of GPS information.

Now let's consider a situation we're outdoors but there are trees above the robot. The problem we have here is actually very similar to the last one. There is a lot of water in the leaves of the trees, and particularly after it has been raining then there's a lot of the water on the surface of the leaves. And this water absorbs the 1.5 gigahertz radio waves from the satellite, and they won't penetrate. So with heavy tree canopy GPS radio waves are absorbed, and we're not going to get enough information for the robot to obtain a fix.

But there's another and perhaps deeper reason why GPS is not the solution to all robotic problems. Consider this kind of typical case where I've got a robot and it wants to pick up an object.

So this robot wants to pick up that banana. Now I can add a GPS receiver to the robot, so the robot now knows where it is in the world. But it doesn't really help, because it doesn't really know where the banana is. So I can add a GPS receiver to the banana, and now the banana knows where it is; the robot knows where it is. But it still hasn't helped with this problem of this robot knowing where the banana is so that it can move and grasp it. So in order to get information about the banana's position to the robot I could add a radio transmitter to this GPS receiver on the banana, and I could add a receiver to the robot. So now the robot knows where the banana is, it has the coordinates of the banana; it knows where it is, and having those two pieces of information, the robot can then plan a path in order to get from here to there.

Now while this might work, the fundamental problem is that every object that the robot would want to pick up, work with and manipulate needs to be fitted with this kind of instrumentation, and clearly this is not the way we solve this problem. If I want to pick up the banana I simply reach out, I use my eyes to work out where it is and guide my hand in order to pick it up.

Robotics Vision

Lecture 1, Section 4

Vision

Let's continue our discussion about the sense of vision. One of the things that makes vision such a powerful sensor is that it is a long-range sensor. It allows us to sense the world beyond our fingertips. Our fingertips, our arms only allow us to reach and explore a relatively small part of the space that is around us. But sometimes the most important things that we need to know about are far beyond the reach of our fingertips; by the time things are that close it is perhaps to late, so vision allows us to see for tens of meters, hundreds of meters, a very great distance and that is a really big virtue of the sense of vision.

Another important characteristic of a sense of vision is that it provides a very rich amount of information. We know something about the shape of the object, we know something about the colour of the object, and that shape and colour in addition to texture allows us to classify the object to work out what sort of thing it is. Is it a thing that we are interested in? Is it something we want to run away from? Or we want to move towards? What sort of object it is, and that is absolutely critical for many, many applications for people and also for robots.

Another great advantage of vision is it gives us information about motion and it tells us about how things are moving around us, so we can see an object and we can see it moving. Is it getting closer to us, is it going further away? But it also provides a lot of information about how we are moving, so as we walk through the world other things appear to move past us due to our own motion.

The fact that so many animals use vision and that the sense of vision has evolved independently multiple times here on planet earth argue, I think, for the effectiveness of the sense of vision. For almost all animals we use the sense of vision for the really important things in life. We use it to help us find food, we use it to avoid being eaten by other animals and we use it to find our mates. So all the important life functions are mediated through the sense of vision.

Have a look at these tasks being preformed by a variety of different animals. Consider the bee, and a bee is a small animal, with a very small brain, but it has got two eyes. Quite complex eyes, very different sorts of eyes to ours, but it can perform quite complicated functions. It can fly for a long distance in order to locate some flowers that other bees have informed it about. It can land on a moving flower or it can tell the difference between a flower and a leaf. So quite complex functions critical to the life of the bee and its colony are performed essentially using vision.

In the middle, we have somebody catching a ball. They are using their eyes to catch the ball. They are not using GPS or any other kind of technological sensor. This person has got their eyes on the ball and that is a very well-known expression. Watching the ball as it moves through the air and they have got some sort of mental model about how balls move through the air. They are using that, then, to plan how their hands should move in order to achieve a successful grasp of the ball. So we are using our eyes to sense the ball. We are making a plan and then we are acting, we are moving our hands in order to intercept it.

And the last example is driving, where we are using our eyes to observe the road, people and cars that are moving on the road. We've got an idea of where it is we want to go. So we are using the visual information with a goal in mind. We are making a plan about how to navigate down this particular bit of road and then we are acting by turning the steering wheel, pushing the brake pedal or the accelerator.

Robotic Vision

Lecture 1, Section 5

Evolution of vision

So life appeared on planet earth about 3.6 billion years ago. But for the bulk of that time, organisms, life on the planet, was not able to see. Around 500 million years ago there was a revolution in the technology of life and the sense of vision appeared, and it appeared very quickly across a number of species and it led to a great diversity in the forms of life. So at this point 500 million years ago called the Cambrian explosion when suddenly all these vision-enabled life forms started to roam the planet.

One of the earliest and perhaps most successful of those life forms is the trilobite, and here I have a fossilised trilobite. They appeared 500 something million years ago and they lived on the planet for something like 270 million years. So they are an animal with a segmented body and they had two fairly primitive, by modern standards, compound eyes. And the sense of vision gave animals a competitive advantage. If an animal could see another animal, perhaps it was a hunting animal and it wanted to prey on that other animal, if it had a sense of vision it had a real advantage, it didn't just have to blindly bump into it. He can actually see where that animal is and go after it and consume it. And similarly if you were a prey animal, the sort of animal that someone might want to eat, then if you had a sense of vision and you had eyes then you could sense a predator coming and you could take some evasive action. So vision led to this… almost an arms race on planet earth as predators and prey animals gain the sense of vision and they improved the sense of vision until we have the sort of very sophisticated vision in very, very many animals today.

So the sense of vision appeared on planet earth about 540 million years ago. And when it appeared it was a game changer. In fact, the sense of vision was so effective that, over time, ten different sorts of eye designs have evolved. For instance, the lensed eye that we have, the compound eye that a insect has, the strange reflector based eyes that scallops have and so on. So there are ten different designs of eyes on the planet. And the lensed eye, like the one we have, has, in fact, been invented seven times across different animal species and across time. So to me these facts argue that vision is a really important and really effective way of sensing what is going on in the world around you.

Now of course vision needs light in order to operate, we need the scene to be illuminated with ambient light. Photons reflected from the scene into our eyes, they fall on the retina and create some neural stimulus. Absolutely has to have a source of light, but of course we have evolved on a planet next to a really bright star. So light is in abundance. It is a very logical solution to the problem. We have an abundant source of light and we have evolved the sensor which has got many advantages, but primarily the ability to sense what is going on at a distance beyond our fingertips.

So how did the sense of vision evolve? There is a lot of speculation about this, but generally it is accepted that light sensitive cells appeared on the bodies of very primitive life forms a long time ago, more than 500 million years ago, and they have the ability to sense brightness or light. And perhaps this was useful to work out which way was up and which way was down and maybe you could see the shadow of something going across and that was a useful thing to have. So these photoreceptors could sense light and they were connected by some sort of primitive nerve fibres to some sort of primitive brain.

This was clearly a useful idea because over time the photoreceptors were arranged into a curve, or in three dimensions, into a cup. And this allowed the photoreceptors to respond to rays that came from a particular direction in space. And this was more useful than just knowing that there was something bright out there, now we knew the direction of the bright thing and this conferred some advantage. This design was refined over time, the chamber became filled with water, perhaps they had better optical properties, and then there was sort of coverings evolved. And then lenses evolved. Until we get to something like the human eye today were we have, you know, protective layers, and we have lenses. We have an iris, which regulates the amount of light which falls onto the retinas. So we can operate quite well in low light and bright light conditions. The photoreceptors, which are this yellow layer shown here, are connected by a bunch of nerve fibres, the optic nerve, which carry those stimuli to the optical or visual parts of the brain. And that is actually the part of the brain located at the back of our head.

Robotic Vision

Lecture 1, Section 6

Principles of vision

The sense of vision has long fascinated human kind.

In the time of the Greeks, they believed that vision was somewhat analogous to the sense of touch. And what happened is when you open your eyes and you looked at an object there was some visual fire, some visual flux left your eye, went out and touched the object and the sense of vision was in fact the interaction between this emanating flux and the object itself. It was like a long distance sense of touch.

Over time the modern interpretation of the way vision works was developed, and certainly by the Middle Ages I think, it was quite well understood. And the way we understand today is that a scene is illuminated by a light source, in this case the sun. Rays of light fall down onto objects and is reflected into all sorts of different directions, and some of those light rays are reflected into our eye, where they are focused onto photoreceptors and lead to some stimulation to the brain, which we interpret as visual perception.

Robotic Vision

Lecture 1, Section 7

Human vision

Let's discuss the human vision system. Now I am not a biologist, so I am going to give you an engineer's impression of how the human vision system works.

Firstly we have the eye itself, and looking at an eye perhaps the most obvious feature is the white part, which is called the sclera, and the iris. Now the sclera is fibrous tissue, which is an important part of the structure of the eye and some kinds of animals the sclera actually has bony components in it and in this fossil here you can see a ring of bone, which is actually part of the sclera of that animal's eye.

If we look at the cross section of the eye on the right here, we can see that the sclera which is part of the membrane that surrounds the eyeball itself. We can see the iris and the pupil which regulates the light going into the eye. We can see the lens, which is a transparent, crystalline protein structure which focuses the light onto the retina on the back of the eye.

The inside of the eye is filled with a fairly transparent fluid, and at the back of the eye we have a surface which is called the retina, a curved surface which contains the photoreceptors, that is, the light sensitive part of the eye. And that is then connected by this big nerve bundle—the optic nerve—to the brain. In fact, we can actually think of the eye as being a part of the brain. It is connected by a significant nerve bundle, and then the eye is essentially a sensory part of the brain located in the front of our heads.

Let's now look in more detail at the retina of the eye. So if we take a transect, we move from inside the eye to outside of the eye along the direction of that arrow we pass through a number of layers of cells and blood vessels and whatever. Now this is all on a pretty small scale; this total structure is about 500 microns thick.

The light sensitive cells are at the bottom of this structure. So they are not on the surface of the retina, they are actually about 500 microns below the surface of the retina. So the light has to pass through—defuse through—these cells and blood vessels before it gets to the light sensitive cells.

Now you've probably heard people talking about rods and cones. The rods cells and the cone cells are the light sensitive elements. They are the photoreceptors in our eye. The rod cells are these spaghetti shaped structures. They are kind of thin. And they are sensitive to low light levels. So we use the rod cells when we are working in a very low light environment. The cone cells are these strange, stumpy shaped cells; they are a little bit conical, kind of squat, not as tall as the rods. And these cells are sensitive to colour. And there are three different types of cones cells: there are some that are sensitive to red light, some sensitive to green light and some sensitive to blue light. But the characteristic they all share is that they are this somewhat conical shape.

Another way we can consider the eye is to look at where the optic fibre joins the retina and that creates an area on the retina which is called the optic disk where there are actually no light sensitive cells. So there is a part of the retina that does not respond to light. And there is another part of the retina where we have an enormous concentration of light sensitive cells and that area is called the fovea. The fovea and the optic nerve are separated by some distance.

So in this graph here what we are going to see is the density of photoreceptor cells, and there is a huge peak in the number of cone cells per square millimetre in this area we call the fovea. So we have exquisite resolution in this one small area of our eye. The rest of the eye, we have much less resolution and unconsciously our eye is continually moving, focusing, on different parts of the scene, directing the fovea to these different parts. So we build up a high resolution of the image by directing the high resolution fovea all over the scene we are looking at.

Now with the rod cells they have a quite different spatial density pattern. There are very few rod cells in the fovea area, but there's many more rod cells in what we call our peripheral vision; away from the direction that our eye is pointing. This is very useful at night and these cells are also somewhat motion sensitive. They are what give us the ability to see something moving out of the corner of our eye.

So what is really interesting then is this business of the optic nerve entering the retina and creating what we call the blind spot. Now we can actually detect our blind spot using a pattern, something like this. So what you do is close your right eye and stare at the plus sign. And then move your head toward and away from the screen making absolutely sure that you are looking with your left eye at the plus sign. As you move your head in and out, the circle here will at some point fall onto the optic disk on your retina and it will, quite surprisingly, it will just disappear from view. So you can have some fun with a test like this, you can find others online. Have a play and see if you can find your blind spot. It is quite surprising that in everyday life we don't notice that we have got a blind spot. We go to a bit of trouble, we can actually detect that it is there.

Now if we look at, for instance, our horizontal cross section of the brain. We can see the eyes located at the front of the head and the optic nerve bundles carry the visual information through the brain to the back part of the brain, where we have our visual cortex. And that is where we have all the really exciting visual processing going on. That is where part of our brain is responsible for recognising people and objects and motion and so on. But the human brain is a very complex structure. It weighs something like 1.5 kilograms and is about 10^{11} neurons in there. So an awful lot of computation and memory capability. Nearly a third of that is devoted to processing visual information.

Earlier we used the argument that the eye has been invented so many times by evolution that it must be a very effective sensor for animals. But there is a cost for the sense of vision, and in the case of human beings one of those costs is that a third of our brain is devoted to processing this visual information. So yes, it is a very effective sensor that as an organism we pay quite a cost in order to run that sensor. Additional costs are associated with the eyes themselves. They are very delicate structures, so we evolved mechanisms to protect the eye. For instance, our eyelids and eyelashes, mechanisms to clean the eyes with tear ducts and so on, and the eyes are also moved within the head by some very high performance muscles. So there is a lot of auxiliary machinery involved in maintaining the sense of vision.

Robotic Vision

Lecture 1, Section 8

Other animals

We have talked a bit about the human visual system. Let's have a look at the way some other animals do it.

Consider the case of the bee. The bee is a very small animal and therefore it has got a very small brain. The bee's brain weighs just one gram and it has only got 10^6 neurons. So contrast that with the human brain, which has got one and a half kilograms of grey matter and 10^{11} neurons. The bee brain is much smaller, much simpler, yet it is still able to do very complex tasks based simply on the information that comes from its eyes.

Now we have two eyes and the bee has two eyes. There is nothing that says that all animals have to have two eyes. So this is a close up of a spider and this particular spider has got four eyes and I think there are some spiders that have got five eyes. So evolution has come up with many different designs of eyes themselves, and configurations of eyes, that clearly confer some particular advantage to the organism that is carrying them.

Here is a close up photograph of the eyes of an insect. These are compound eyes, made up of a number of polygonal facets.

Here is another type of compound eye. This one doesn't have quite the field of view of the one we just looked at.

Here is the eye of a nautilus, which is quite a primitive ocean-going animal. And the eye really is a very simple pinhole camera; it doesn't have a lens as such.

Here are the eyes of a scallop. So the blue objects that you can see, each of those is a very simple eye and it has got a large number of them around the outside of what is effectively its mouth. Each of these eyes is a bit like a reflecting telescope. The inside of those blue objects is quite shiny and so incoming light rays, instead of refracting in through a lens, reflect off the shinny surface inside that eye and fall onto the photoreceptors. So this scallop, it uses reflection rather then refraction and it has got a very large number of eyes.

This is the eye of a squid and it has got a lensed eye like we have, and it has evolved completely independently of the lensed eye that we have.

And here is a close up of the eyes of a bird.

The compound eye in the insect contains a number of facets; individual light sensing elements called ommatidia and if we look at a cross-section of the ommatidia it has got effectively a lens on top which funnels any light that falls on it down inside this conical cell which contains the light sensitive chemicals, which omit a neural pulse when light falls on it. So each one of these is effectively a single pixel camera, or we can think of it in those terms. And we have a whole array of them pointing in slightly different directions and that allows the insect to create a fairly large field of view.

Robotic Vision

Lecture 1, Section 9

Vision as a process

The way we see the world is actually quite complex. Our eyes are not like a camera that takes an image that we then go and process. As we look at a scene our eye moves continuously to different points in the scene in order to improve its understanding of what is in front of it. So seeing is actually a very active process.

So researchers have developed technology that allows us to study where a human being is looking; involves a camera which looks at the subject's eyes, and perhaps an illumination system. So with this we can work out continuously and in real time where a human being is looking.

Now this painting was used in a very famous study that was conducted back in 1967 into how a human subject's gaze depends on the particular question that they were asked. So what they could do is take a subject, ask them a question and then track where their eye looked within the painting. So here is an example: here is one trace, one subject trace. You can see that their gaze is moving all around the room. So of these three questions, which question do you think was asked that led to this particular gaze pattern by that human subject. I will let you think about that for a moment.

And here is the answer. The question that was asked is: 'What are the material circumstances of the family?'.

So in order to answer this question, the human subject has directed their gaze towards the possessions of the particular family. So they're checking out the furniture, they are checking out the paintings on the wall and so on. So given a specific question to a subject, our eye moves over the scene in a particular way so as to best answer the question. It's not just a matter of taking a picture and processing it: our eye actively moves around the scene in order to best answer the question.

Here is another experiment. See if you can work out which question has been asked.

We can see that the gaze is checking out particularly the facial regions of the particular people here. So the question that was asked is: 'What are the ages of the figures in the painting?'. So clearly it is useful to look at the faces of the people in order to work out what their age is.

Third one, in this particular case this is what the subject's gaze did. We can see him checking out the faces, and it is also checking out the whole body of each of the human subjects in this painting. And the question that was asked is: 'What type of clothes are the family wearing?'.

So given the subject was answering that particular question, then their gaze is focused just on the faces and the clothes of each of the people. It is not looking at the furniture; it is not looking at the paintings; it's not looking out the window.

It is important to understand that seeing is a very active process. Now it is driven by perhaps some of the fastest moving muscles in the human body. So the human eye is able to rotate at up to six hundred degrees per second, and has a phenomenal acceleration, something like 35000º per second squared, so amazing muscles are able to point the eye very quickly anywhere within the field of view of the eyes.

So this lovely quote that I like very much by a very famous early vision researcher David Marr. And he says vision is 'the process of discovering from images what is present in the world and where it is'. So vision is a process. I think that is a very important message to take away when we start to think about how robots might see.

Robotic Vision

Lecture 1, Section 10

Robot vision

We have covered a lot of topics. We have talked about how the sense of vision evolved. We have talked about the details of human vision. We have looked at some of the variety of animal vision. We have talked about how vision is a process where the eye is actively steered around the scene in order to maximise our understanding of that particular scene.

So now it is appropriate to talk about robots and vision, and look at how robots have used the sense of vision in order to be able to perform tasks, in a similar way to which humans and animals use vision to perform tasks.

The robot on the left is a very famous robot called Shakey. Developed at SRI International in the 1960s. And it has a television camera on the top, which it uses for navigation. The robot on the right is one that I built in the early 2000s, and it has got quite a number of cameras on board, and they are a little bit interesting to look at. In the front is a stereo camera pair and it is a bit like our own two eyes. It allows the robot to determine three dimensional structure of the world in front of it. On the top, it has a very shiny mirror object and that is part of what we call a panoramic camera assembly. There is a camera looking up at that mirror and that allows the robot to see 360 degrees—it can see forwards, backwards and side ways. It's sometimes called an omni-cam or a panoramic camera.

The last camera on this robot is a wide-angle camera on the side. It has a fish eye lens, so it almost has a hemispherical field of view. And that can look at what is happening to the side of the robot.

The current Mars Rover, the Curiosity Rover, has got a large number of cameras on board that it uses for a variety of functions.

Now the business of teaching machines to see is the field of computer vision, and that is an activity that has been going on since perhaps the 1960s. Some important early work was carried out by Larry Roberts at MIT as part of his PhD project. Larry went on to do other very important things and was very instrumental in creating the internet. So in his theses work what he was doing was taking a picture of an object, and this is a very simple wooden block object, took a picture with a TV camera and was trying to work out what shape it was. So he went through a number of vision processing steps and these are the sorts of things we will cover in following lectures.

One of the first steps was to find the edges of this object and once he had found the edges of the object he would try to fit line segments to those edges and once he knew those line segments, he knew something about the way images are formed. He could then say something about the three dimensional shape of this particular object. So what we have here is a system that can take an image and process it and come up with a three dimensional model of that shape. So it is a very simple form of object recognition. The computers take an image and made a discussion about what sort of object that it was looking at. That's sufficient information then for a robot to make a move, go and pick that object up and perhaps manipulate it in some way.

So why is vision a good sensor for a robot to have? There are a few reasons why I believe vision is an important and very practical sensor for a robot.

Firstly, the cameras themselves are now very cheap and the reason for this is because cameras are built into everything; built into cell phones, built into laptops, and so on. So the actual sensor the equivalent of a retina is now a device that perhaps costs less then a dollar. Lenses are smaller, cameras are small and cheap.

The other reason that is really important is that computation is now really cheap; we have very powerful computer chips and lots and lots of memory, and so this enables us to run algorithms to process the data that comes out of the sensor chip.

So this combination of very effective, high resolution, colour, cheap sensors, with abundant computation are the foundations on which we can build robot vision systems.

Here is a really interesting graph from Ray Kurzweil, and he talks a lot about the way computation power has changed over time, and this is a logarithmic vertical scale and his plotting a number of data points that represent a number of calculations per second you can buy for a thousand dollars over time. And we can see that this is an exponential plot on a logarithmic vertical scale. So what it is showing is computation is really increasing, increasingly rapidly with time. And this is fundamentally Ray Kurzweil's theses.

So if we extrapolate this into the future, we can see that we are about here, and we have computational power of effectively one mouse brain.

By the early 2020s we should have, for a thousand dollars, the computational power of a human brain. And by 2050 when many of you will be alive, perhaps at the ends of your working careers, for a thousand dollars you will be able to buy enough computing power to the equivalent of all human brains on the planet. That is a pretty amazing prediction and so clearly very, very exciting times ahead.

So what practical things have roboticists been doing with sensor vision?

We hear a lot in recent times about self-driving cars, Google cars and so on, but the actual history of self-driving cars goes back a long way. There was some very significant research program in Europe in the 1980s called Prometheus, and a lot of very fundamental work was done by a scientist called Ernst Dickmanns. He automated this van. It had a number of cameras in the front looking outwards, and it was able to drive along autobahns at high speed. A number of significant landmark achievements were made by this particular van and some of it immediate descendants.

Another landmark achievement was from some researchers from Carnegie Mellon University, who automated a car primarily using the sensor vision, and drove it across America. The journey took a few days, around four thousand kilometres, and there were relatively few human interventions.

This video shows a humanoid robot catching a ball. Here we see it again in slow motion. Now the robot has a pair of cameras in its head, which allows it to estimate the distance to the ball; uses that information to model how a ball moves through space in order to plan the motion that the arm should take to intercept the ball. So the robot's head has a number of sensors: pair of cameras as I mentioned, but also some tilt sensors, so it can work out where the head is pointing in space. Now you can see the positions of the ball as seen by its left and right eye as a function of time, and here you can see an animation of the robot's hand moving to intercept the path of the ball.

Here is a flying robot which we actually looked at up close in the *Out and About with Robots* video, in the Robotics course some of you might have seen earlier. This robot is equipped with a stereo pair of cameras again. So again like our own two eyes, this enables the robot to sense an obstacle in front of it, by working out the three dimensional structure of the world in front using information from two cameras and a fair amount of processing on board.

Another robot developed by myself and some colleagues at CSIRO, and this underwater robot has also got a stereo pair of cameras. It has got two cameras that look downwards and two cameras which look frontwards. The downward looking cameras are estimating the distance of the seabed from the robot, and this robot is trying to maintain a constant altitude above the sea bed. And it does that using the three dimensional information from the downward looking stereo cameras. The frontward looking stereo cameras are used to detect obstacles.

Here we see what the world looks like through the eyes of a mobile robot. This particular robot has got a stereo camera pair which allows it to create a three dimensional model of the world through which it is moving. And that is really useful in order to determine what is a flat surface that it could drive over, and what is a wall, or human being, or some other kind of obstacle.

Here we have something a bit different. Here we have a single camera looking downwards at a coral reef, and the camera is being carried by a robot. Now we are able to use a number of mathematical techniques to combine the information from these multiple camera videos to create a three dimensional model of the coral reef. We do this from a number of single camera views. Now we smooth that three dimensional mesh; we drape the original imagery over it to create a texture map surface. So now we have a very realistic looking three dimensional model of a coral reef obtained just from a whole sequence of single camera views.

So I hope that I have convinced you that vision is a really, really important sensor for all sorts of animals, for ourselves and also for robots. So in the rest of the course we are going to learn something about how we do robot vision; how do we take information from camera sensors and process it and generate information that a robot can take some action on.